

01-21-00

A

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (4/98)
Approved for use through 09/30/2000 OMB 0651-0032
Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 C.F.R. § 1.53(b))

Attorney Docket No. 600-1-195B

First Inventor or Application Identifier James E. Darnell, Jr.

Title NUCLEIC ACIDS ENCODING RECEPTOR ...

Express Mail Label No. EL406398184US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

1. ☒ * Fee Transmittal Form (e.g., PTO/SB/17)
(Submit an original and a duplicate for fee processing)
2. ☒ Specification [Total Pages 107]
(preferred arrangement set forth below)
- Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 42]
4. Oath or Declaration [Total Pages 3]
- a. ☐ Newly executed (original or copy)
 - b. ☒ Copy from a prior application (37 C.F.R. § 1.63(d))
(for continuation/divisional with Box 16 completed)
 - i. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).

5. ☐ Microfiche Computer Program (Appendix)
6. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
- a. ☐ Computer Readable Copy
 - b. ☒ Paper Copy (identical to computer copy)
 - c. ☒ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

7. ☒ Assignment Papers (cover sheet & document(s))
8. ☐ 37 C.F.R. § 3.73(b) Statement (when there is an assignee) ☒ Power of Attorney
9. ☐ English Translation Document (if applicable)
10. ☒ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
11. ☒ Preliminary Amendment
12. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
- * Small Entity ☐ Statement filed in prior application, Status still proper and desired (PTO/SB/09-12)
13. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)
14. ☐ Other: Petition for Filing Date and Decision
15. ☒ Other: Petition for Filing Date and Decision

NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).

16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:

☒ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No. 08/948,547

Prior application information Examiner L. SPECTOR Group / Art Unit 1646

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or ☐ Correspondence address below

Name David A. Jackson
Klauber & Jackson

Address Continental Plaza
411 Hackensack Avenue

City Hackensack State New Jersey Zip Code 07601

Country U.S.A. Telephone 1(201)487-5800 Fax 1(201)343-1684

Name (Print/Type) Michael D. Davis Registration No. (Attorney/Agent) 39,161

Signature Michael D. Davis Date 1/20/2000

Burden Hour Statement This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: JAMES E. DARNELL, JR. ET AL.

SERIAL NO. : UNASSIGNED EXAMINER : UNKNOWN

FILED : HEREWITH ART UNIT : UNKNOWN

FOR : NUCLEIC ACIDS ENCODING RECEPTOR RECOGNITION
FACTORS, AND METHODS OF USE THEREOF (AMENDED)

VIA EXPRESS MAIL: EL 406398184 US
DATE OF DEPOSIT: JANUARY 24 2000

PRELIMINARY AMENDMENT

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

Dear Sir:

In accordance with Rule 111 of the Rules of Practice please consider the following
amendments and remarks.

IN THE SPECIFICATION:

Page 1, line 6, after "Application" insert -- is a Continuation of copending U.S. Serial
No. 08/948,547, filed October 10, 1997, which is a Division of copending U.S. Serial No.
08/212,185, filed March 11, 1994 which --; on line 8, replace "1994" with -- 1993, both now
abandoned, --; on line 9, after "1992" insert --, now abandoned, --; and line 10, after "1992"
insert --, now abandoned, --.

On Page 5, line 27 after the open parenthesis "(" please insert -- *i.e.*, the murine
homologue of the human protein having an amino acid sequence of --.

On Page 7, line 10 replace "translation protein" with -- transcription factor --.

On Page 9, lines 10, 15, 20, and 25 please replace "sequece" with -- sequence --.

On Page 17, line 23 after "FIGURE" please replace "1" with - - 1A-1E - - ; and
on line 28 after "FIGURE" please replace "2" with - - 2A-2D - -.

On Page 18, line 1 after "FIGURE" please replace "3" with - - 3A-3C - -; and
on line 19 replace "5" with -- 5a-5b --.

Page 19, line 4 replace "7" with -- 7a-7e --.

Page 23, line 5, before "the DNA", replace "(B-D)" with -- (B-C) --; line 9, before
"the DNA", replace "(B-C)" with -- (B-D) --; and line 13, before "the DNA", replace "(B-C)"
with -- (B-E) --.

On Page 38, lines 4, 10, 14, and 19 please replace "sequece" with -- sequence --.

Page 70, line 4, before "and SEQ ID NO:7", replace "12A-12C" with -- 13A-13C --.

Page 76, line 11, after "EXAMPLE" insert -- 6 --.

Applicants request that the Specification be amended to include the Sequence Listing submitted herewith at the end of the Specification and prior to the Claims. Applicants enclose a copy of the Sequence Listing for the Examiner's convenience. Applicants request that the Specification be renumbered as follows: The Sequence Listing should now be on numbered pages 93-134. The pages containing the Claims as filed should be renumbered from pages 93- 106 to pages 135-148. The page containing the abstract should be renumbered from page 107 to page 149.

Please amend the title of the application to read:

NUCLEIC ACIDS ENCODING RECEPTOR RECOGNITION FACTORS AND
METHODS OF USE THEREOF.

IN THE CLAIMS:

Please cancel Claims 2-68 without prejudice:

Please add the following new claims.

--69. A recombinant DNA molecule encoding a receptor recognition factor (RRF) protein having the following characteristics:

- a) said RRF is cytoplasmic in origin;
- b) said RRF is activated by tyrosine phosphorylation;
- c) upon activation said RRF is translocated to the nucleus of a target cell; and
- d) said RRF has an amino acid sequence comprising a sequence of contiguous amino acid residues which is present in both SEQ ID NO:2 and SEQ ID NO:4; wherein the sequence of contiguous amino acid residues contains four or more consecutive amino acids.

70. The recombinant DNA molecule of Claim 69 wherein the sequence of contiguous amino acid residues contains four or more consecutive amino acids and is selected from the group consisting of:

- a) HQLY (amino acids 19-22 of SEQ ID NO:2 and 19-22 of SEQ ID NO4);
- b) IRQY (amino acids 31-34 of SEQ ID NO:2 and 30-33 of SEQ ID NO4);
- c) RQYL (amino acids 32-35 of SEQ ID NO:2 and 31-34 of SEQ ID NO4);
- d) LLQH (amino acids 82-85 of SEQ ID NO:2 and 78-81 of SEQ ID NO4);
- e) LQHN (amino acids 83-86 of SEQ ID NO:2 and 79-82 of SEQ ID NO4);
- f) RKEV (amino acids 210-213 of SEQ ID NO:2 and 210-213 of SEQ ID NO4);
- g) FVVE (amino acids 316-319 of SEQ ID NO:2 and 317-320 of SEQ ID NO4);
- h) QPCM (amino acids 321-324 of SEQ ID NO:2 and 322-325 of SEQ ID NO4);
- i) PCMP (amino acids 322-325 of SEQ ID NO:2 and 323-326 of SEQ ID NO4);
- j) LKTG (amino acids 334-337 of SEQ ID NO:2 and 335-338 of SEQ ID NO4);

k) RLLV (amino acids 345-348 of SEQ ID NO:2 and 346-349 of SEQ ID NO:4);
l) GFRK (amino acids 372-375 of SEQ ID NO:2 and 376-379 of SEQ ID NO:4);
m) FRKF (amino acids 373-376 of SEQ ID NO:2 and 377-380 of SEQ ID NO:4);
n) RKFN (amino acids 374-377 of SEQ ID NO:2 and 378-381 of SEQ ID NO:4);
o) KFNI (amino acids 375-378 of SEQ ID NO:2 and 379-382 of SEQ ID NO:4);
p) FNIL (amino acids 376-379 of SEQ ID NO:2 and 380-383 of SEQ ID NO:4);
q) VTEE (amino acids 424-427 of SEQ ID NO:2 and 426-429 of SEQ ID NO:4);
r) TEEL (amino acids 425-428 of SEQ ID NO:2 and 427-430 of SEQ ID NO:4);
s) EELH (amino acids 426-429 of SEQ ID NO:2 and 428-431 of SEQ ID NO:4);
t) LPVV (amino acids 451-454 of SEQ ID NO:2 and 453-456 of SEQ ID NO:4);
u) LSWQ (amino acids 500-503 of SEQ ID NO:2 and 502-505 of SEQ ID NO:4);
v) SWQF (amino acids 501-504 of SEQ ID NO:2 and 503-506 of SEQ ID NO:4);
w) WQFS (amino acids 502-505 of SEQ ID NO:2 and 504-507 of SEQ ID NO:4);
x) QFSS (amino acids 503-506 of SEQ ID NO:2 and 505-508 of SEQ ID NO:4);
y) RGLN (amino acids 510-513 of SEQ ID NO:2 and 512-515 of SEQ ID NO:4);
z) ILEL (amino acids 560-563 of SEQ ID NO:2 and 561-564 of SEQ ID NO:4);
zz) LWND (amino acids 571-574 of SEQ ID NO:2 and 572-575 of SEQ ID NO:4);
aa) WNDG (amino acids 572-575 of SEQ ID NO:2 and 573-576 of SEQ ID NO:4);
bb) IMGF (amino acids 577-580 of SEQ ID NO:2 and 578-581 of SEQ ID NO:4);
cc) GTFL (amino acids 596-599 of SEQ ID NO:2 and 597-600 of SEQ ID NO:4);
dd) TFLI (amino acids 597-600 of SEQ ID NO:2 and 598-601 of SEQ ID NO:4);
ee) FLLR (amino acids 598-601 of SEQ ID NO:2 and 599-602 of SEQ ID NO:4);
ff) LLRF (amino acids 599-602 of SEQ ID NO:2 and 600-603 of SEQ ID NO:4);
gg) LRFS (amino acids 600-603 of SEQ ID NO:2 and 601-604 of SEQ ID NO:4);
hh) RFSE (amino acids 601-604 of SEQ ID NO:2 and 602-605 of SEQ ID NO:4);
ii) FSES (amino acids 602-605 of SEQ ID NO:2 and 603-606 of SEQ ID NO:4);
jj) SESS (amino acids 603-606 of SEQ ID NO:2 and 604-607 of SEQ ID NO:4);
kk) PYTK (amino acids 630-633 of SEQ ID NO:2 and 633-636 of SEQ ID NO:4);
ll) ENIP (amino acids 654-657 of SEQ ID NO:2 and 657-660 of SEQ ID NO:4);
mm) NIPE (amino acids 655-658 of SEQ ID NO:2 and 658-661 of SEQ ID NO:4);

nn) IPEN (amino acids 656-659 of SEQ ID NO:2 and 659-662 of SEQ ID NO:4);
oo) PENP (amino acids 657-660 of SEQ ID NO:2 and 660-663 of SEQ ID NO:4); and
pp) ENPL (amino acids 658-661 of SEQ ID NO:2 and 661-664 of SEQ ID NO:4).

71. The recombinant DNA molecule of Claim 70 wherein the sequence of contiguous amino acid residues contains five or more consecutive amino acids and is selected from the group consisting of:

- a) IRQYL (amino acids 31-35 of SEQ ID NO:2 and 30-34 of SEQ ID NO:4);
- b) LLQHN (amino acids 82-86 of SEQ ID NO:2 and 78-82 of SEQ ID NO:4);
- c) QPCMP (amino acids 321-325 of SEQ ID NO:2 and 322-326 of SEQ ID NO:4);
- d) GFRKF (amino acids 372-376 of SEQ ID NO:2 and 376-380 of SEQ ID NO:4);
- e) FRKFN (amino acids 373-377 of SEQ ID NO:2 and 377-381 of SEQ ID NO:4);
- f) RKFNI (amino acids 374-378 of SEQ ID NO:2 and 378-382 of SEQ ID NO:4);
- g) KFNIL (amino acids 375-379 of SEQ ID NO:2 and 379-383 of SEQ ID NO:4);
- h) VTEEL (amino acids 424-428 of SEQ ID NO:2 and 426-430 of SEQ ID NO:4);
- i) TEELH (amino acids 425-429 of SEQ ID NO:2 and 427-431 of SEQ ID NO:4);
- j) LSWQF (amino acids 500-504 of SEQ ID NO:2 and 502-506 of SEQ ID NO:4);
- k) SWQFS (amino acids 501-505 of SEQ ID NO:2 and 503-507 of SEQ ID NO:4);
- l) WQFSS (amino acids 502-506 of SEQ ID NO:2 and 504-508 of SEQ ID NO:4);
- m) LWNDG (amino acids 571-575 of SEQ ID NO:2 and 572-576 of SEQ ID NO:4);
- n) GTFLL (amino acids 596-600 of SEQ ID NO:2 and 597-601 of SEQ ID NO:4);
- o) TFLLR (amino acids 597-601 of SEQ ID NO:2 and 598-602 of SEQ ID NO:4);
- p) FLLRF (amino acids 598-602 of SEQ ID NO:2 and 599-603 of SEQ ID NO:4);
- q) LLRFS (amino acids 599-603 of SEQ ID NO:2 and 600-604 of SEQ ID NO:4);
- r) LRFSE (amino acids 600-604 of SEQ ID NO:2 and 601-605 of SEQ ID NO:4);
- s) RFSES (amino acids 601-605 of SEQ ID NO:2 and 602-606 of SEQ ID NO:4);
- t) FSESS (amino acids 602-606 of SEQ ID NO:2 and 603-607 of SEQ ID NO:4);
- u) ENIPE (amino acids 654-658 of SEQ ID NO:2 and 657-661 of SEQ ID NO:4);
- v) NIPEN (amino acids 655-659 of SEQ ID NO:2 and 658-662 of SEQ ID NO:4);

w) IPENP (amino acids 656-660 of SEQ ID NO:2 and 659-663 of SEQ ID NO:4);
and

x) PENPL (amino acids 657-661 of SEQ ID NO:2 and 660-664 of SEQ ID NO:4).

72. The recombinant DNA molecule of Claim 71 wherein the sequence of contiguous amino acid residues contains six or more consecutive amino acids and is selected from the group consisting of:

- a) GFRKFN (amino acids 372-377 of SEQ ID NO:2 and 376-381 of SEQ ID NO:4);
- b) FRKFNI (amino acids 373-378 of SEQ ID NO:2 and 377-382 of SEQ ID NO:4);
- c) RKFNIL (amino acids 374-379 of SEQ ID NO:2 and 378-383 of SEQ ID NO:4);
- d) VTEELH (amino acids 424-429 of SEQ ID NO:2 and 426-431 of SEQ ID NO:4);
- e) LSWQFS (amino acids 500-505 of SEQ ID NO:2 and 502-507 of SEQ ID NO:4);
- f) SWQFSS (amino acids 501-506 of SEQ ID NO:2 and 503-508 of SEQ ID NO:4);
- g) GTFLLR (amino acids 596-601 of SEQ ID NO:2 and 597-602 of SEQ ID NO:4);
- h) TFLRLF (amino acids 597-602 of SEQ ID NO:2 and 598-603 of SEQ ID NO:4);
- i) FLLRFS (amino acids 598-603 of SEQ ID NO:2 and 599-604 of SEQ ID NO:4);
- j) LLRFSE (amino acids 599-604 of SEQ ID NO:2 and 600-605 of SEQ ID NO:4);
- k) LRFSES (amino acids 600-605 of SEQ ID NO:2 and 601-606 of SEQ ID NO:4);
- l) RFSESS (amino acids 601-606 of SEQ ID NO:2 and 602-607 of SEQ ID NO:4);
- m) ENIPEN (amino acids 654-659 of SEQ ID NO:2 and 657-662 of SEQ ID NO:4);
- n) NIPENP (amino acids 655-660 of SEQ ID NO:2 and 658-663 of SEQ ID NO:4);

and

o) IPENPL (amino acids 656-661 of SEQ ID NO:2 and 659-664 of SEQ ID NO:4).

73. The recombinant DNA molecule of Claim 72 wherein the sequence of contiguous amino acid residues contains seven or more consecutive amino acids and is selected from the group consisting of:

- a) GFRKFNI (amino acids 372-378 of SEQ ID NO:2 and 376-382 of SEQ ID NO:4);
- b) FRKFNIL (amino acids 373-379 of SEQ ID NO:2 and 377-383 of SEQ ID NO:4);
- c) LSWQFSS (amino acids 500-506 of SEQ ID NO:2 and 502-508 of SEQ ID NO:4);

- d) GTFLLRF (amino acids 596-602 of SEQ ID NO:2 and 597-603 of SEQ ID NO:4);
- e) TFLLRFS (amino acids 597-603 of SEQ ID NO:2 and 598-604 of SEQ ID NO:4);
- f) FLLRFSE (amino acids 598-604 of SEQ ID NO:2 and 599-605 of SEQ ID NO:4);
- g) LLRFSES (amino acids 599-605 of SEQ ID NO:2 and 600-606 of SEQ ID NO:4);
- h) LRFSESS (amino acids 600-606 of SEQ ID NO:2 and 601-607 of SEQ ID NO:4);
- i) ENIPENP (amino acids 654-660 of SEQ ID NO:2 and 657-663 of SEQ ID NO:4);

and

- j) NIPENPL (amino acids 655-661 of SEQ ID NO:2 and 658-664 of SEQ ID NO:4).

74. The recombinant DNA molecule of Claim 73 wherein the sequence of contiguous amino acid residues contains eight or more consecutive amino acids and is selected from the group consisting of:

- a) GFRKFNIL (amino acids 372-379 of SEQ ID NO:2 and 376-383 of SEQ ID NO:4);
- b) GTFLLRFS (amino acids 596-603 of SEQ ID NO:2 and 597-604 of SEQ ID NO:4);
- c) TFLLRFS (amino acids 597-604 of SEQ ID NO:2 and 598-605 of SEQ ID NO:4);
- d) FLLRFSES (amino acids 598-605 of SEQ ID NO:2 and 599-606 of SEQ ID NO:4);
- e) LLRFSESS (amino acids 599-606 of SEQ ID NO:2 and 600-607 of SEQ ID NO:4); and
- f) ENIPENPL (amino acids 654-661 of SEQ ID NO:2 and 657-664 of SEQ ID NO:4).

75. The recombinant DNA molecule of Claim 74 wherein the sequence of contiguous amino acid residues contains nine or more consecutive amino acids and is selected from the group consisting of:

- a) GTFLLRFSE (amino acids 596-604 of SEQ ID NO:2 and 597-605 of SEQ ID NO:4);

b) TFLLRFSES (amino acids 597-605 of SEQ ID NO:2 and 598-606 of SEQ ID NO:4); and

c) FLLRFSESS (amino acids 598-606 of SEQ ID NO:2 and 599-607 of SEQ ID NO:4).

76. The recombinant DNA molecule of Claim 75 wherein the sequence of contiguous amino acid residues contains ten or more consecutive amino acids and is selected from the group consisting of:

a) GTFLLRFSES (amino acids 596-605 of SEQ ID NO:2 and 597-606 of SEQ ID NO:4); and

b) TFLLRFSESS (amino acids 597-606 of SEQ ID NO:2 and 598-607 of SEQ ID NO:4).

77. The recombinant DNA molecule of Claim 76 wherein the sequence of contiguous amino acid residues contains eleven consecutive amino acids having the sequence GTFLLRFSESS (amino acids 596-606 of SEQ ID NO:2 and 597-607 of SEQ ID NO:4).

78. The recombinant DNA molecule of Claim 70 wherein said RRF has an amino acid sequence which further comprises a second sequence of contiguous amino acid residues, wherein the second sequence of contiguous amino acid residues also contains four or more consecutive amino acids which is present in both SEQ ID NO:2 and SEQ ID NO:4.

79. A recombinant DNA molecule encoding a receptor recognition factor (RRF) protein having the following characteristics:

a) said RRF is cytoplasmic in origin;

b) said RRF is activated by tyrosine phosphorylation; and

c) upon activation said RRF is translocated to the nucleus of a target cell, wherein said DNA molecule hybridizes to the nucleotide sequence set forth in SEQ ID NO:1 under standard hybridization conditions.

80. A recombinant DNA molecule encoding a receptor recognition factor (RRF) protein having the following characteristics:

- a) said RRF is cytoplasmic in origin;
- b) said RRF is activated by tyrosine phosphorylation; and
- c) upon activation said RRF is translocated to the nucleus of a target cell; wherein said DNA molecule hybridizes to the nucleotide sequence set forth in SEQ ID NO:3 under standard hybridization conditions.

81. A recombinant DNA molecule encoding a receptor recognition factor (RRF) protein having the following characteristics:

- (a) the RRF is cytoplasmic in origin;
- (b) the RRF is activated by tyrosine phosphorylation; and
- (c) upon activation said RRF is translocated to the nucleus of a target cell;

wherein the RRF contains one or more of the boxed regions in Figure 8B.

82 The recombinant DNA molecule of Claim 81, wherein the RRF further contains a tyrosyl residue at a position that corresponds to the conserved position identified in SEQ ID NO:2 and SEQ ID NO:4, said position selected from the group consisting of:

- amino acid 22 of SEQ ID NO:2 and amino acid 22 of SEQ ID NO:4;
- amino acid 34 of SEQ ID NO:2 and amino acid 33 of SEQ ID NO:4;
- amino acid 288 of SEQ ID NO:2 and amino acid 289 of SEQ ID NO:4;
- amino acid 631 of SEQ ID NO:2 and amino acid 634 of SEQ ID NO:4;
- amino acid 648 of SEQ ID NO:2 and amino acid 651 of SEQ ID NO:4;
- amino acid 665 of SEQ ID NO:2 and amino acid 668 of SEQ ID NO:4;
- amino acid 677 of SEQ ID NO:2 and amino acid 680 of SEQ ID NO:4;
- amino acid 678 of SEQ ID NO:2 and amino acid 681 of SEQ ID NO:4; and
- amino acid 690 of SEQ ID NO:2 and amino acid 701 of SEQ ID NO:4.

83. The recombinant DNA molecule of Claim 81 wherein the RRF comprises a highly negative charged domain at its C-terminal end.
84. The recombinant DNA molecule of Claim 81 wherein the RRF comprises an SH2 domain.
85. The recombinant DNA molecule of Claim 84 wherein the SH2 domain contains an arginine at a position that corresponds to amino acid 601 of SEQ ID NO:2 and amino acid 602 of SEQ ID NO:4.
86. The recombinant DNA molecule of Claim 81 wherein the RRF forms a dimer upon said activation by tyrosine phosphorylation.
87. The recombinant DNA molecule of Claim 81 wherein the activation of the RRF is unaffected by the presence or concentration of second messengers.
88. The recombinant DNA molecule of Claim 81 wherein the RRF can act as a DNA binding protein upon said activation by tyrosine phosphorylation.
89. The recombinant DNA molecule of Claim 81 wherein the RRF interacts with an interferon- γ -bound receptor kinase complex.
90. The recombinant DNA molecule of Claim 88 wherein the RRF can stimulate ISRE-dependent or gamma activated site (GAS)-dependent transcription.
91. An isolated nucleic acid encoding a receptor recognition factor (RRF) protein having the following characteristics:
- (a) the RRF is cytoplasmic in origin;
 - (b) the RRF is activated by tyrosine phosphorylation; and

(c) upon activation said RRF is translocated to the nucleus of a target cell; wherein the RRF contains one or more of the boxed regions in Figure 8B.

92. The isolated nucleic acid of Claim 91, wherein the RRF further contains a tyrosyl residue at a position that corresponds to the conserved position identified in SEQ ID NO:2 and SEQ ID NO:4, said position selected from the group consisting of:

amino acid 22 of SEQ ID NO:2 and amino acid 22 of SEQ ID NO:4;
amino acid 34 of SEQ ID NO:2 and amino acid 33 of SEQ ID NO:4;
amino acid 288 of SEQ ID NO:2 and amino acid 289 of SEQ ID NO:4;
amino acid 631 of SEQ ID NO:2 and amino acid 634 of SEQ ID NO:4;
amino acid 648 of SEQ ID NO:2 and amino acid 651 of SEQ ID NO:4;
amino acid 665 of SEQ ID NO:2 and amino acid 668 of SEQ ID NO:4;
amino acid 677 of SEQ ID NO:2 and amino acid 680 of SEQ ID NO:4;
amino acid 678 of SEQ ID NO:2 and amino acid 681 of SEQ ID NO:4; and
amino acid 690 of SEQ ID NO:2 and amino acid 701 of SEQ ID NO:4.

93. The recombinant DNA molecule of Claim 81 that is operatively linked to an expression control sequence.

94. An expression vector containing the recombinant DNA molecule of Claim 93.

95. A method of expressing a recombinant receptor recognition factor in a cell containing the expression vector of Claim 94 comprising culturing the cell in an appropriate cell culture medium under conditions that provide for expression of the receptor recognition factor by the cell.

96. The method of Claim 95 further comprising the step of purifying the recombinant receptor recognition factor.--

REMARKS

The Specification has been amended to incorporate references to the appropriate SEQ ID NOs:, where specific sequences are indicated. Applicants have amended the Specification to note the status of the parent applications, above and have also addressed the issues concerning 37 CFR 1.821(c). The Specification has been amended as described above, in order to correct obvious typographical errors.

At the outset, Applicants bring to the Examiner's attention the NOTICE OF INCOMPLETE APPLICATION issued on 4/20/94 in the co-pending parent application, Serial No. 08/212,185. Applicants responded on 5/26/94 with a PETITION FOR FILING DATE UNDER 37 C.F.R. § 1.53(b) which was favorably received in the DECISION ON PETITION issued by the Special Program Examiner on August 8, 1994. Copies of these papers are included in the present filing.

Support for the newly added claims can be found throughout the present Specification as indicated below, and more specifically in the Specification as originally filed on March 19, 1992, *see* lines 1-21 on Page 37 of the present Specification. Support for Claims 69-80 as related to the properties of the claimed RRFs may be found on lines 18-29 of Page 4; on lines 1-12 of Page 8; on lines 13-24 of Page 12; and throughout the first three Examples. Further support to the claimed sequence homology between SEQ ID NOs:2 and 4 may be found on lines 8-10, and 15-22 on Page 6; on lines 3-5 of Page 20; and throughout Figure 8b. Further support for Claims 79-80 may be found on lines 15-18 of Page 6; Claims 14 and 15; lines 25-27 on Page 32; on lines 6-8 of Page 35; and throughout the first three Examples. Further support for Claims 81 and 99 may be found in the Specification on line 18 of Page 4 through line 3 of Page 5, on lines 1-12 of Page 8, on lines 13-24 of Page 12, on lines 3-10 of Page 21, and in Figure 8B, on lines 10-30 of Page 35, and throughout the Examples. Further support for Claims 82 and 92 may be found on lines 8-10 of Page 21, and in Figure 8B of the Specification. Further support for Claim 83 may be found on lines 28-30 of Page 20 of the Specification. Further support for Claims 84-86 may be found in the Specification on lines 1-8 of Page 5, on lines 22-29 of Page 7, on lines 4-18 of Page 26, line 4 of Page 83 through line 27 of Page 87, and in Figures 19-23. Further support for Claim 87 may be found on

Page 5, lines 10-14, and Page 7, lines 1-2 of the Specification. Further support for Claims 88-90 may be found in the Specification on Page 7, lines 9-20, on Page 8, lines 1-12, on Page 24, lines 23-29, on Page 36, lines 7-24, and in Figure 18. Further support for Claims 93-96 may be found in the Specification on lines 2-7, and 14-26, of Page 10, on lines 5-12 of Page 11, and on lines 15-18 of Page 16. Claims 1, and 69-96 remain for consideration.

No additional fees are believed to be necessitated by the foregoing amendments. However, should this be erroneous, authorization is hereby given to charge Deposit Account No. 11-1153 for any underpayment, or credit any overages.

Applicants respectfully request entry of the foregoing amendment into the file history of the above-identified Application being filled herewith. Early and favorable action on the pending set of Claims is earnestly solicited.

Respectfully submitted,



MICHAEL D. DAVIS
Attorney for Applicant(s)
Registration No. 39,161

KLAUBER & JACKSON
411 Hackensack Avenue
Hackensack, New Jersey 07601
(201) 487-5800
Date: January 20, 2000

**RECEPTOR RECOGNITION FACTORS, PROTEIN SEQUENCES
AND METHODS OF USE THEREOF**

CROSS-REFERENCE TO RELATED APPLICATIONS

5

The present Application is a Continuation-In-Part of copending U.S. Serial No. 08/126,588 and copending U.S. Serial No. 08/126,595, both filed September 24, 1994, which are both Continuations-In-Part of copending U.S. Serial No. 07/980,498, filed November 23, 1992, which is a Continuation-In-Part of
10 copending U.S. Serial No. 07/854,296, filed March 19, 1992, the disclosures of which are hereby incorporated by reference in their entireties. Applicants claim the benefits of these Applications under 35 U.S.C. § 120.

RELATED PUBLICATIONS

15

The Applicants are authors or co-authors of several articles directed to the subject matter of the present invention. (1) Darnell et al., "Interferon-Dependent Transcriptional Activation: Signal Transduction Without Second Messenger Involvement?" THE NEW BIOLOGIST, 2(10):1-4, (1990); (2) X. Fu et al.,
20 "ISGF3, The Transcriptional Activator Induced by Interferon α , Consists of Multiple Interacting Polypeptide Chains" PROC. NATL. ACAD. SCI. USA, 87:8555-8559 (1990); (3) D.S. Kessler et al., "IFN α Regulates Nuclear Translocation and DNA-Binding Affinity of ISGF3, A Multimeric Transcriptional Activator" GENES AND DEVELOPMENT, 4:1753 (1990). All of the above
25 listed articles are incorporated herein by reference.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to intracellular receptor recognition
30 proteins or factors(i.e. groups of proteins), and to methods and compositions including such factors or the antibodies reactive toward them, or analogs thereof in assays and for diagnosing, preventing and/or treating cellular debilitation, derangement or dysfunction. More particularly, the present invention relates to

particular IFN-dependent receptor recognition molecules that have been identified and sequenced, and that demonstrate direct participation in intracellular events, extending from interaction with the liganded receptor at the cell surface to transcription in the nucleus, and to antibodies or to other entities specific thereto
 5 that may thereby selectively modulate such activity in mammalian cells.

BACKGROUND OF THE INVENTION

There are several possible pathways of signal transduction that might be followed
 10 after a polypeptide ligand binds to its cognate cell surface receptor. Within minutes of such ligand-receptor interaction, genes that were previously quiescent are rapidly transcribed (Murdoch et al., 1982; Larner et al., 1984; Friedman et al., 1984; Greenberg and Ziff, 1984; Greenberg et al., 1985). One of the most physiologically important, yet poorly understood, aspects of these immediate
 15 transcriptional responses is their specificity: the set of genes activated, for example, by platelet-derived growth factor (PDGF), does not completely overlap with the one activated by nerve growth factor (NGF) or tumor necrosis factor (TNF) (Cochran et al., 1983; Greenberg et al., 1985; Almendral et al., 1988; Lee et al., 1990). The interferons (IFN) activate sets of other genes entirely. Even
 20 IFN α and IFN γ , whose presence results in the slowing of cell growth and in an increased resistance to viruses (Tamm et al., 1987) do not activate exactly the same set of genes (Larner et al., 1984; Friedman et al., 1984; Celis et al., 1987, 1985; Larner et al., 1986).

25 The current hypotheses related to signal transduction pathways in the cytoplasm do not adequately explain the high degree of specificity observed in polypeptide-dependent transcriptional responses. The most commonly discussed pathways of signal transduction that might ultimately lead to the nucleus depend on properties of cell surface receptors containing tyrosine kinase domains [for example, PDGF, epidermal growth factor (EGF), colony-stimulating factor (CSF), insulin-like
 30 growth factor-1 (IGF-1); see Gill, 1990; Hunter, 1990) or of receptors that interact

with G-proteins (Gilman, 1987). These two groups of receptors mediate changes in the intracellular concentrations of second messengers that, in turn, activate one of a series of protein phosphokinases, resulting in a cascade of phosphorylations (or dephosphorylations) of cytoplasmic proteins.

5

It has been widely conjectured that the cascade of phosphorylations secondary to changes in intracellular second messenger levels is responsible for variations in the rates of transcription of particular genes (Bourne, 1988, 1990; Berridge, 1987; Gill, 1990; Hunter, 1990). However, there are at least two reasons to question the suggestion that global changes in second messengers participate in the chain of events leading to specific transcriptional responses dependent on specific receptor occupation by polypeptide ligands.

First, there is a limited number of second messengers (cAMP, diacyl glycerol, phosphoinositides, and Ca^{2+} are the most prominently discussed), whereas the number of known cell surface receptor-ligand pairs of only the tyrosine kinase and G-protein varieties, for example, already greatly outnumbers the list of second messengers, and could easily stretch into the hundreds (Gill, 1990; Hunter, 1990). In addition, since many different receptors can coexist on one cell type at any instant, a cell can be called upon to respond simultaneously to two or more different ligands with an individually specific transcriptional response each involving a different set of target genes. Second, a number of receptors for polypeptide ligands are now known that have neither tyrosine kinase domains nor any structure suggesting interaction with G-proteins. These include the receptors for interleukin-2 (IL-2) (Leonard et al., 1985), $\text{IFN}\alpha$ (Uze et al., 1990), $\text{IFN}\gamma$ (Aguet et al., 1988), NGF (Johnson et al., 1986), and growth hormone (Leung et al., 1987). The binding of each of these receptors to its specific ligand has been demonstrated to stimulate transcription of a specific set of genes. For these reasons it seems unlikely that global intracellular fluctuations in a limited set of second messengers are integral to the pathway of specific, polypeptide ligand-dependent, immediate transcriptional responses.

In PCT International Publication No. WO 92/08740 published 29 May, 1992 by the applicant herein, the above analysis was presented and it was discovered and proposed that a receptor recognition factor or factors, served in some capacity as a type of direct messenger between liganded receptors at the cell surface and the cell nucleus. One of the characteristics that was ascribed to the receptor recognition factor was its apparent lack of requirement for changes in second messenger concentrations. Continued investigation of the receptor recognition factor through study of the actions of the interferons IFN α and IFN γ has further elucidated the characteristics and structure of the interferon-related factor ISGF-3, and more broadly, the characterization and structure of the receptor recognition factor in a manner that extends beyond earlier discoveries previously described. It is accordingly to the presentation of this updated characterization of the receptor recognition factor and the materials and methods both diagnostic and therapeutic corresponding thereto that the present disclosure is directed.

15

SUMMARY OF THE INVENTION

In accordance with the present invention, receptor recognition factors have been further characterized that appear to interact directly with receptors that have been occupied by their ligand on cellular surfaces, and which in turn either become active transcription factors, or activate or directly associate with transcription factors that enter the cells' nucleus and specifically binds on predetermined sites and thereby activates the genes. It should be noted that the receptor recognition proteins thus possess multiple properties, among them: 1) recognizing and being activated during such recognition by receptors; 2) being translocated to the nucleus by an inhibitable process (eg. NaF inhibits translocation); and 3) combining with transcription activating proteins or acting themselves as transcription activation proteins, and that all of these properties are possessed by the proteins described herein.

30

A further property of the receptor recognition factors (also termed herein signal transducers and activators of transcription -- STAT) is dimerization to form homodimers or heterodimers upon activation by phosphorylation of tyrosine. In a specific embodiment, *infra*, Stat91 and Stat84 form homodimers and a Stat91-

- 5 Stat84 heterodimer. Accordingly, the present invention is directed to such dimers, which can form spontaneously by phosphorylation of the STAT protein, or which can be prepared synthetically by chemically cross-linking two like or unlike STAT proteins.
- 10 The receptor recognition factor is proteinaceous in composition and is believed to be present in the cytoplasm. The recognition factor is not demonstrably affected by concentrations of second messengers, however does exhibit direct interaction with tyrosine kinase domains, although it exhibits no apparent interaction with G-proteins. More particularly, as is shown in a co-pending, co-owned application
- 15 entitled "INTERFERON-ASSOCIATED RECEPTOR RECOGNITION FACTORS, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF USE THEREOF," filed on even date herewith, the 91 kD human interferon (IFN) γ factor, represented by SEQ ID NO:4 directly interacts with DNA after acquiring phosphate on tyrosine located at position 701 of the amino acid
- 20 sequence.

The recognition factor is now known to comprise several proteinaceous substituents, in the instance of IFN α and IFN γ . Particularly, three proteins derived from the factor ISGF-3 have been successfully sequenced and their

- 25 sequences are set forth in FIGURE 1 (SEQ ID NOS:1, 2), FIGURE 2 (SEQ ID NOS:3, 4) and FIGURE 3 (SEQ. ID NOS.5, 6) herein. Additionally, a murine gene encoding the 91 kD protein (SEQ ID NO:4) has been identified and sequenced. The nucleotide sequence (SEQ ID NO:7) and deduced amino acid sequence (SEQ ID NO:8) are shown in FIGURE 13A-13C.

In a further embodiment, murine genes encoding homologs of the recognition factor have been successfully sequenced and cloned into plasmids. A gene in plasmid 13sf1 has the nucleotide sequence (SEQ ID NO:9) and deduced amino acid sequence (SEQ ID NO:10) as shown in FIGURE 14A-14C. A gene in
5 plasmid 19sf6 has the nucleotide sequence (SEQ ID NO:11) and deduced amino acid sequence (SEQ ID NO:12) shown in FIGURE 15A-15C.

It is particularly noteworthy that the protein sequence of FIGURE 1 (SEQ ID NO:2) and the sequence of the proteins of FIGURES 2 (SEQ ID NO:4) and 3
10 (SEQ ID NO:6) derive, respectively, from two different but related genes. Moreover, the protein sequence of FIGURE 13 (SEQ ID NO:8) derives from a murine gene that is analogous to the gene encoding the protein of FIGURE 2 (SEQ ID NO:4). Of further note is that the protein sequences of FIGURES 14 (SEQ ID NO:10) and 15 (SEQ ID NO:12) derive from two genes that are different from,
15 but related to, the protein of FIGURE 13 (SEQ ID NO:8). It is clear from these discoveries that a family of genes exists, and that further family members likewise exist. Accordingly, as demonstrated herein, by use of hybridization techniques, additional such family members will be found.

20 Further, the capacity of such family members to function in the manner of the receptor recognition factors disclosed, herein may be assessed by determining those ligand that cause the phosphorylation of the particular family members.

In its broadest aspect, the present invention extends to a receptor recognition
25 factor implicated in the transcriptional stimulation of genes in target cells in response to the binding of a specific polypeptide ligand to its cellular receptor on said target cell, said receptor recognition factor having the following characteristics:

- a) apparent direct interaction with the ligand-bound receptor complex
30 and activation of one or more transcription factors capable of binding with a specific gene;

- b) an activity demonstrably unaffected by the presence or concentration of second messengers;
- c) direct interaction with tyrosine kinase domains; and
- d) a perceived absence of interaction with G-proteins.

5

In a further aspect, the receptor recognition (STAT) protein forms a dimer upon activation by phosphorylation.

In a specific example, the receptor recognition factor represented by SEQ ID NO:4 possesses the added capability of acting as a translation protein and, in particular, as a DNA binding protein in response to interferon- γ stimulation. This discovery presages an expanded role for the proteins in question, and other proteins and like factors that have heretofore been characterized as receptor recognition factors. It is therefore apparent that a single factor may indeed provide the nexus between the liganded receptor at the cell surface and direct participation in DNA transcriptional activity in the nucleus. This pleiotypic factor has the following characteristics:

- a) It interacts with an interferon- γ -bound receptor kinase complex;
- b) It is a tyrosine kinase substrate; and
- c) When phosphorylated, it serves as a DNA binding protein.

More particularly, the factor represented by SEQ ID NO:4 is interferon-dependent in its activity and is responsive to interferon stimulation, particularly that of interferon- γ . It has further been discovered that activation of the factor represented by SEQ ID NO:4 requires phosphorylation of tyrosine-701 of the protein, and further still that tyrosine phosphorylation requires the presence of a functionally active SH2 domain in the protein. Preferably, such SH2 domain contains an amino acid residue corresponding to an arginine at position 602 of the protein.

30

In a still further aspect, the present invention extends to a receptor recognition factor interactive with a liganded interferon receptor, which receptor recognition factor possesses the following characteristics:

- a) it is present in cytoplasm;
- 5 b) it undergoes tyrosine phosphorylation upon treatment of cells with IFN α or IFN γ ;
- c) it activates transcription of an interferon stimulated gene;
- d) it stimulates either an ISRE-dependent or a gamma activated site (GAS)-dependent transcription *in vivo*;
- 10 e) it interacts with IFN cellular receptors, and
- f) it undergoes nuclear translocation upon stimulation of the IFN cellular receptors with IFN.

The factor of the invention represented by SEQ ID NO:4 appears to act in similar
15 fashion to an earlier determined site-specific DNA binding protein that is interferon- γ dependent and that has been earlier called the γ activating factor (GAF). Specifically, interferon- γ -dependent activation of this factor occurs without new protein synthesis and appears within minutes of interferon- γ treatment, achieves maximum extent between 15 and 30 minutes thereafter, and
20 then disappears after 2-3 hours. These further characteristics of identification and action assist in the evaluation of the present factor for applications having both diagnostic and therapeutic significance.

In a particular embodiment, the present invention relates to all members of the
25 herein disclosed family of receptor recognition factors except the 91 kD protein factors, specifically the proteins whose sequences are represented by one or more of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

The present invention also relates to a recombinant DNA molecule or cloned gene,
30 or a degenerate variant thereof, which encodes a receptor recognition factor, or a fragment thereof, that possesses a molecular weight of about 113 kD and an amino

acid sequence set forth in FIGURE 1 (SEQ ID NO:2); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 113 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA sequence shown in FIGURE 1 (SEQ ID NO:1). In another

5 embodiment, the receptor recognition factor has a molecular weight of about 91 kD and the amino acid sequence set forth in FIGURE 2 (SEQ ID NO:4) or FIGURE 13 (SEQ ID NO:8); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 91 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA

10 seqeence shown in FIGURE 2 (SEQ ID NO:3) or FIGURE 13 (SEQ ID NO:8). In yet a further embodiment, the receptor recognition factor has a molecular weight of about 84 kD and the amino acid sequence set forth in FIGURE 3 (SEQ ID NO:6); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 84 kD receptor recognition factor has a

15 nucleotide sequence or is complementary to a DNA seqeence shown in FIGURE 3 (SEQ ID NO:5). In yet another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIGURE 14 (SEQ ID NO:10); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is

20 complementary to a DNA seqeence shown in FIGURE 14 (SEQ ID NO:9). In still another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIGURE 15 (SEQ ID NO:12); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA

25 seqeence shown in FIGURE 15 (SEQ ID NO:11).

The human and murine DNA sequences of the receptor recognition factors of the present invention or portions thereof, may be prepared as probes to screen for complementary sequences and genomic clones in the same or alternate species.

30 The present invention extends to probes so prepared that may be provided for screening cDNA and genomic libraries for the receptor recognition factors. For

example, the probes may be prepared with a variety of known vectors, such as the phage λ vector. The present invention also includes the preparation of plasmids including such vectors, and the use of the DNA sequences to construct vectors expressing antisense RNA or ribozymes which would attack the mRNAs of any or
5 all of the DNA sequences set forth in FIGURES 1, 2, 3, 13, 14 and 15 (SEQ ID NOS:1, 3, 5, 7, 9, and 11, respectively). Correspondingly, the preparation of antisense RNA and ribozymes are included herein.

The present invention also includes receptor recognition factor proteins having the
10 activities noted herein, and that display the amino acid sequences set forth and described above and selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10 and SEQ ID NO:12.

In a further embodiment of the invention, the full DNA sequence of the
15 recombinant DNA molecule or cloned gene so determined may be operatively linked to an expression control sequence which may be introduced into an appropriate host. The invention accordingly extends to unicellular hosts transformed with the cloned gene or recombinant DNA molecule comprising a DNA sequence encoding the present receptor recognition factor(s), and more
20 particularly, the complete DNA sequence determined from the sequences set forth above and in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9 and SEQ ID NO:11.

According to other preferred features of certain preferred embodiments of the
25 present invention, a recombinant expression system is provided to produce biologically active animal or human receptor recognition factor.

The concept of the receptor recognition factor contemplates that specific factors exist for correspondingly specific ligands, such as tumor necrosis factor, nerve
30 growth factor and the like, as described earlier. Accordingly, the exact structure of each receptor recognition factor will understandably vary so as to achieve this

ligand and activity specificity. It is this specificity and the direct involvement of the receptor recognition factor in the chain of events leading to gene activation, that offers the promise of a broad spectrum of diagnostic and therapeutic utilities.

- 5 The present invention naturally contemplates several means for preparation of the recognition factor, including as illustrated herein known recombinant techniques, and the invention is accordingly intended to cover such synthetic preparations within its scope. The isolation of the cDNA amino acid sequences disclosed herein facilitates the reproduction of the recognition factor by such recombinant
10 techniques, and accordingly, the invention extends to expression vectors prepared from the disclosed DNA sequences for expression in host systems by recombinant DNA techniques, and to the resulting transformed hosts.

- The invention includes an assay system for screening of potential drugs effective to
15 modulate transcriptional activity of target mammalian cells by interrupting or potentiating the recognition factor or factors. In one instance, the test drug could be administered to a cellular sample with the ligand that activates the receptor recognition factor, or an extract containing the activated recognition factor, to determine its effect upon the binding activity of the recognition factor to any
20 chemical sample (including DNA), or to the test drug, by comparison with a control.

- The assay system could more importantly be adapted to identify drugs or other entities that are capable of binding to the receptor recognition and/or transcription
25 factors or proteins, either in the cytoplasm or in the nucleus, thereby inhibiting or potentiating transcriptional activity. Such assay would be useful in the development of drugs that would be specific against particular cellular activity, or that would potentiate such activity, in time or in level of activity. For example, such drugs might be used to modulate cellular response to shock, or to treat other
30 pathologies, as for example, in making IFN more potent against cancer.

In yet a further embodiment, the invention contemplates antagonists of the activity of a receptor recognition factor (STAT). In particular, an agent or molecule that inhibits dimerization (homodimerization or heterodimerization) can be used to block transcription activation effected by an activated, phosphorylated STAT protein. In a specific embodiment, the antagonist can be a peptide having the sequence of a portion of an SH2 domain of a STAT protein, or the phosphotyrosine domain of a STAT protein, or both. If the peptide contains both regions, preferably the regions are located in tandem, more preferably with the SH2 domain portion N-terminal to the phosphotyrosine portion. In a specific example, *infra*, such peptides are shown to be capable of disrupting dimerization of STAT proteins.

One of the characteristics of the present receptor recognition factors is their participation in rapid phosphorylation and dephosphorylation during the course of and as part of their activity. Significantly, such phosphorylation takes place in an interferon-dependent manner and within a few minutes in the case of the ISGF-3 proteins identified herein, on the tyrosine residues defined thereon. This is strong evidence that the receptor recognition factors disclosed herein are the first true substrates whose intracellular function is well understood and whose intracellular activity depends on tyrosine kinase phosphorylation. In particular, the addition of phosphate to the tyrosine of a transcription factor is novel. This suggests further that tyrosine kinase takes direct action in the transmission of intracellular signals to the nucleus, and does not merely serve as a promoter or mediator of serine and/or serine kinase activity, as has been theorized to date. Also, the role of the factor represented by SEQ ID NO:2 in its activated phosphorylated form suggests possible independent therapeutic use for this activated form. Likewise, the role of the factor as a tyrosine kinase substrate suggests its interaction with kinase in other theatres apart from the complex observed herein.

The diagnostic utility of the present invention extends to the use of the present receptor recognition factors in assays to screen for tyrosine kinase inhibitors.

Because the activity of the receptor recognition-transcriptional activation proteins described herein must maintain tyrosine phosphorylation, they can and presumably are dephosphorylated by specific tyrosine phosphatases. Blocking of the specific phosphatase is therefore an avenue of pharmacological intervention that would
5 potentiate the activity of the receptor recognition proteins.

The present invention likewise extends to the development of antibodies against the receptor recognition factor(s), including naturally raised and recombinantly prepared antibodies. For example, the antibodies could be used to screen
10 expression libraries to obtain the gene or genes that encode the receptor recognition factor(s). Such antibodies could include both polyclonal and monoclonal antibodies prepared by known genetic techniques, as well as bi-specific (chimeric) antibodies, and antibodies including other functionalities suiting them for additional diagnostic use conjunctive with their capability of modulating
15 transcriptional activity.

In particular, antibodies against specifically phosphorylated factors can be selected and are included within the scope of the present invention for their particular ability in following activated protein. Thus, activity of the recognition factors or
20 of the specific polypeptides believed to be causally connected thereto may therefore be followed directly by the assay techniques discussed later on, through the use of an appropriately labeled quantity of the recognition factor or antibodies or analogs thereof.

25 Thus, the receptor recognition factors, their analogs and/or analogs, and any antagonists or antibodies that may be raised thereto, are capable of use in connection with various diagnostic techniques, including immunoassays, such as a radioimmunoassay, using for example, an antibody to the receptor recognition factor that has been labeled by either radioactive addition, reduction with sodium
30 borohydride, or radioiodination.

In an immunoassay, a control quantity of the antagonists or antibodies thereto, or the like may be prepared and labeled with an enzyme, a specific binding partner and/or a radioactive element, and may then be introduced into a cellular sample. After the labeled material or its binding partner(s) has had an opportunity to react
5 with sites within the sample, the resulting mass may be examined by known techniques, which may vary with the nature of the label attached. For example, antibodies against specifically phosphorylated factors may be selected and appropriately employed in the exemplary assay protocol, for the purpose of following activated protein as described above.

10

In the instance where a radioactive label, such as the isotopes ^3H , ^{14}C , ^{32}P , ^{35}S , ^{36}Cl , ^{51}Cr , ^{57}Co , ^{58}Co , ^{59}Fe , ^{90}Y , ^{125}I , ^{131}I , and ^{186}Re are used, known currently available counting procedures may be utilized. In the instance where the label is an enzyme, detection may be accomplished by any of the presently utilized
15 colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques known in the art.

The present invention includes an assay system which may be prepared in the form of a test kit for the quantitative analysis of the extent of the presence of the
20 recognition factors, or to identify drugs or other agents that may mimic or block their activity. The system or test kit may comprise a labeled component prepared by one of the radioactive and/or enzymatic techniques discussed herein, coupling a label to the recognition factors, their agonists and/or antagonists, and one or more additional immunochemical reagents, at least one of which is a free or
25 immobilized ligand, capable either of binding with the labeled component, its binding partner, one of the components to be determined or their binding partner(s).

In a further embodiment, the present invention relates to certain therapeutic
30 methods which would be based upon the activity of the recognition factor(s), its (or their) subunits, or active fragments thereof, or upon agents or other drugs

determined to possess the same activity. A first therapeutic method is associated with the prevention of the manifestations of conditions causally related to or following from the binding activity of the recognition factor or its subunits, and comprises administering an agent capable of modulating the production and/or activity of the recognition factor or subunits thereof, either individually or in mixture with each other in an amount effective to prevent the development of those conditions in the host. For example, drugs or other binding partners to the receptor recognition/transcription factors or proteins may be administered to inhibit or potentiate transcriptional activity, as in the potentiation of interferon in cancer therapy. Also, the blockade of the action of specific tyrosine phosphatases in the dephosphorylation of activated (phosphorylated) recognition/transcription factors or proteins presents a method for potentiating the activity of the receptor recognition factor or protein that would concomitantly potentiate therapies based on receptor recognition factor/protein activation.

15

More specifically, the therapeutic method generally referred to herein could include the method for the treatment of various pathologies or other cellular dysfunctions and derangements by the administration of pharmaceutical compositions that may comprise effective inhibitors or enhancers of activation of the recognition factor or its subunits, or other equally effective drugs developed for instance by a drug screening assay prepared and used in accordance with a further aspect of the present invention. For example, drugs or other binding partners to the receptor recognition/transcription factor or proteins, as represented by SEQ ID NO:2, may be administered to inhibit or potentiate transcriptional activity, as in the potentiation of interferon in cancer therapy. Also, the blockade of the action of specific tyrosine phosphatases in the dephosphorylation of activated (phosphorylated) recognition/transcription factor or protein presents a method for potentiating the activity of the receptor recognition factor or protein that would concomitantly potentiate therapies based on receptor recognition factor/protein activation. Correspondingly, the inhibition or blockade of the activation or binding of the recognition/transcription factor would affect MHC

Class II expression and consequently, would promote immunosuppression. Materials exhibiting this activity, as illustrated later on herein by staurosporine, may be useful in instances such as the treatment of autoimmune diseases and graft rejection, where a degree of immunosuppression is desirable.

5

In particular, the proteins of ISGF-3 whose sequences are presented in SEQ ID NOS:2, 4, 6, 8, 10 or 12 herein, their antibodies, agonists, antagonists, or active fragments thereof, could be prepared in pharmaceutical formulations for administration in instances wherein interferon therapy is appropriate, such as to
10 treat chronic viral hepatitis, hairy cell leukemia, and for use of interferon in adjuvant therapy. The specificity of the receptor proteins hereof would make it possible to better manage the aftereffects of current interferon therapy, and would thereby make it possible to apply interferon as a general antiviral agent.

15 Accordingly, it is a principal object of the present invention to provide a receptor recognition factor and its subunits in purified form that exhibits certain characteristics and activities associated with transcriptional promotion of cellular activity.

20 It is a further object of the present invention to provide antibodies to the receptor recognition factor and its subunits, and methods for their preparation, including recombinant means.

It is a further object of the present invention to provide a method for detecting the
25 presence of the receptor recognition factor and its subunits in mammals in which invasive, spontaneous, or idiopathic pathological states are suspected to be present.

It is a further object of the present invention to provide a method and associated assay system for screening substances such as drugs, agents and the like,
30 potentially effective in either mimicking the activity or combating the adverse effects of the recognition factor and/or its subunits in mammals.

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the recognition factor or subunits thereof, so as to alter the adverse consequences of such presence or activity, or where beneficial, to enhance such activity.

5

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the recognition factor or its subunits, so as to treat or avert the adverse consequences of invasive, spontaneous or idiopathic pathological states.

10

It is a still further object of the present invention to provide pharmaceutical compositions for use in therapeutic methods which comprise or are based upon the recognition factor, its subunits, their binding partner(s), or upon agents or drugs that control the production, or that mimic or antagonize the activities of the

15 recognition factors.

Other objects and advantages will become apparent to those skilled in the art from a review of the ensuing description which proceeds with reference to the following illustrative drawings.

20

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3 α gene defining the 113 kD protein. The nucleotides are numbered from 1 to 2553 (SEQ ID NO:1), and the amino acids are numbered from 1 to 851 (SEQ ID NO:2).

FIGURE 2 depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3 α gene defining the 91 kD protein. The nucleotides are numbered from 1 to 3943 (SEQ ID NO:3), and the amino acids are numbered from 1 to 750 (SEQ ID NO:4).

FIGURE 3 depicts the full receptor recognition factor nucleic acid sequence and the deduced amino acid sequence derived for the ISGF-3 α gene defining the 84 kD protein. The nucleotides are numbered from 1 to 2166 (SEQ ID NO:5), and the amino acids are numbered from 1 to 712 (SEQ ID NO:6).

5

FIGURE 4 shows the purification of ISGF-3. The left-hand portion of the Figure shows the purification of ISGF-3 demonstrating the polypeptides present after the first oligonucleotide affinity column (lane 3) and two different preparations after the final chromatography step (Lanes 1 and 2). The left most lane contains
 10 protein size markers (High molecular weight, Sigma). ISGF-3 component proteins are indicated as 113 kD, 91 kD, 84 kD, and 48 kD [Kessler et al., *GENES & DEV.*, 4 (1990); Levy et al., *THE EMBO. J.*, 9 (1990)]. The right-hand portion of the Figure shows purified ISGF-3 from $2-3 \times 10^{11}$ cells was electroblotted to nitrocellulose after preparations 1 and 2 (Lanes 1 and 2) had been pooled and
 15 separated on a 7.5% SDS polyacrylamide gel. ISGF-3 component proteins are indicated. The two lanes on the right represent protein markers (High molecular weight, and prestained markers, Sigma).

FIGURE 5 generally presents the results of Northern Blot analysis for the 91/84
 20 kD peptides. Figure 5a presents restriction maps for cDNA clones E4 (top map) and E3 (bottom map) showing DNA fragments that were radiolabeled as probes (probes A-D). Figure 5b comprises Northern blots of cytoplasmic HeLa RNA hybridized with the indicated probes. The 4.4 and 3.1 KB species as well as the 28S and 18S rRNA bands are indicated.

25

FIGURE 6 depicts the conjoint protein sequence of the 91 kD (SEQ ID NO:4) and 84 kD (SEQ ID NO:6) proteins of ISGF-3. One letter amino acid code is shown for the open reading frame from clone E4, (encoding the 91 kD protein). The 84 kD protein, encoded by a different cDNA (E3), has the identical sequence but
 30 terminates after amino acid 712, as indicated. Tryptic peptides t19, t13a, and t13b

from the 91 kD protein are indicated. The sole recovered tryptic peptide from the 84 kD protein, peptide t27, was wholly contained within peptide t19 as indicated.

FIGURE 7 presents the results of Western blot and antibody shift analyses.

- 5 a) Highly purified ISGF-3, fractionated on a 7.0% SDS polyacrylamide gel, was probed with antibodies a42 (amino acids 597-703); a55 (amino acids 2-59); and a57 (amino acids 705-739) in a Western blot analysis. The silver stained part of the gel (lanes a, b, and c) illustrates the location of the ISGF-3 component proteins and the purity of the material used in Western blot: Lane a) 10 Silver stain of protein sample used in all the Western blot experiments (immune and preimmune). Lane b) Material of equal purity to that shown in Fig. 4, for clearer identification of the ISGF-3 proteins. Lane c) Size protein markers indicated.
- 15 b) Antibody interference of the ISGF-3 shift complex; Lane a) The complete ISGF-3 and the free ISGF-3 γ component shift with partially purified ISGF-3 are marked; Lane b) Competition with a 100 fold excess of cold ISRE oligonucleotide. Lane c) Shift complex after the addition of 1 ml of preimmune serum to a 12.5 μ l shift reaction. Lanes d and e) - Shift complex after the addition of 1 μ l of a 1:10 dilution or 1 ml of undiluted a42 antiserum to a 12.5 μ l shift 20 reaction.

Methods:

- Antibodies a42, a55 and a57 were prepared by injecting approximately 500 mgm of a fusion protein prepared in *E. coli* using the GE3-3X vector [Smith et al., 25 *GENE*, 67 (1988)]. Rabbits were bled after the second boost and serum prepared.

- For Western blots highly purified ISGF-3 was separated on a 7% SDS polyacrylamide gel and electroblotted to nitrocellulose. The filter was incubated in blocking buffer ("blotto"), cut into strips and probed with specific antiserum and 30 preimmune antiserum diluted 1:500. The immune complexes were visualized with the aid of an ECL kit (Amersham). Shift analyses were performed as previously

described [Levy et al., *GENES & DEV.*, 2 (1988); Levy et al., *GENES & DEV.*, 3 (1989)] in a 4.5% polyacrylamide gel.

FIGURE 8 presents the full length amino acid sequence of 113 kD protein components of ISGF-3 α (SEQ ID NO:2) and alignment of conserved amino acid sequences between the 113 kD and 91/84 kD proteins (SEQ ID NOS:4 AND 6).

A. Polypeptide sequences (A-E) derived from protein micro-sequencing of purified 113 kD protein (see accompanying paper) are underlined. Based on peptide E, we designed a degenerate oligonucleotide,

10 AAT/CACIGAA/GCCATGGAA/GATT/CATT (SEQ ID NO:13), which was used to screen a cDNA library [Pine et al., *MOL. CELL. BIOL.*, 10 (1990)] basically as described [Norman et al., *CELL*, 55 (1988)]. Briefly, the degenerate oligonucleotides were labeled by ^{32}P - γ -ATP by polynucleotide kinase, hybridizations were carried out overnight at 40°C in 6 x SSTE (0.9 M NaCl, 60 mM Tris-HCl [pH 7.9] 6mM EDTA), 0.1% SDS, 2mM Na₂P₅O₇, 6 mM KH₂PO₄, 15 in the presence of 100 mg/ml salmon sperm DNA and 10 x Denhardt's solution [Maniatis et al., *MOLECULAR CLONING; A LABORATORY MANUAL* (Cold Spring Harbor Lab., 1982)]. The nitrocellulose filters then were washed 4 x 10 min. with the same hybridization conditions without labeled probe and 20 salmon sperm DNA. Autoradiography was carried out at -80°C with intensifying screen for 48 hrs. A PCR product was obtained later by the same method described for the 91/84 kD sequences, by using oligonucleotides designed according polypeptide D and E. The sequence of this PCR product was identical to a region in clone f11. The full length of 113 kD protein contains 851 amino 25 acids. Three major helices in the N-terminal region were predicted by the methods of both Chou and Fasman [Chou et al., *ANN. REV. BIOCHEM.*, 47 (1978)] and Garnier et al [Garnier et al., *J. MOL. BIOL.*, 12 (1978)] and are shown in shadowed boxes. At the C-terminal end, a highly negative charged domain was found. All negative charged residues are blackened and positive 30 charged residues shadowed. The five polypeptides that derived from protein

microscreening [Aebersold et al., *PROC. NATL. ACAD. SCI. USA*, **87** (1987)] are underlined.

- B) Comparison of amino acid sequences of 113 kD and 91/84 kD protein shows a 42% identical amino acid residues in the overlapping 715 amino acid sequence shown. In the middle helix region four leucine and one valine heptad repeats were identified in both 113 and 91/84 kD protein (the last leucine in 91/84 kD is not exactly preserved as heptad repeats). When a heligram structure was drawn this helix is amphipathic (not shown). Another notable feature of this comparison is several tyrosine residues that are conserved in both proteins near their ends.

FIGURE 9 shows the *in vitro* transcription and translation of 113 kD and 91 kD cDNA and a Northern blot analysis with 113 kD cDNA probe.

- a) The full length cDNA clones of 113 and 91 kD protein were transcribed *in vitro* and transcribed RNAs was translated *in vitro* with rabbit lenticulate lysate (Promega; conditions as described in the Promega protocol). The mRNA of BMV (Promega) was simultaneously translated as a protein size marker. The 113 cDNA yielded a translated product about 105 kD and the 91 cDNA yielded a 86 kD product.
- b) When total cytoplasmic mRNAs isolated from superinduced HeLa cells were utilized, a single 4.8 KB mRNA band was observed with a cDNA probe coding for C-end of 113 kD protein in a Northern blot analysis [Nielsch et al., *The EMBO. J.*, **10** (1991)].
- FIGURE 10(A) presents the results of Western blot analysis confirming the identity of the 113 kD protein. An antiserum raised against a polypeptide segment [Harlow et al., *ANTIBODIES; A LABORATORY MANUAL* (Cold Spring Harbor Lab., 1988)] from amino acid 500 to 650 of 113 kD protein recognized specifically a 113 kD protein in a protein Western blot analysis. The antiserum recognized a band both in a highly purified ISGF-3 fraction (> 10,000 fold) from DNA affinity chromatography and in the crude extracts prepared from γ and α

IFN treated HeLa cells [Fu et al., *PROC. NATL. ACAD. SCI. USA*, **87** (1990)]. The antiserum was raised against a fusion protein [a cDNA fragment coding for part of 113 kD protein was inserted into pGEX-2T, a high expression vector in the E. coli [Smith et al., *PROC. NATL. ACAD. SCI. USA*, **83** (1986)] purified from

5 E. coli [Smith et al., *GENE*, **67** (1988)]. The female NZW rabbits were immunized with 1 mg fusion protein in Freund's adjuvant. Two subsequent boosts two weeks apart were carried out with 500 mg fusion protein. The Western blot was carried out with conditions described previously [Pine et al., *MOL. CELL. BIOL.*, **10** (1990)].

10

FIGURE 10(B) presents the results of a mobility shift assay showing that the anti-113 antiserum affects the ISGF-3 shift complex. Preimmune serum or the 113 kD antiserum was added to shift reaction carried out as described [Fu et al. *PROC. NATL. ACAD. SCI. USA*, **87** (1990); Kessler et al. *GENES & DEV.*, **4**,

15 (1990)] at room temperature for 20 min. then one-third of reaction material was loaded onto a 5% polyacrylamide gel. In addition unlabeled probe was included in one reaction to show specificity of the gel shift complexes.

FIGURE 11 shows the results of experiments investigating the IFN- α dependent

20 phosphorylation of 113, 91 and 84 kD proteins. Protein samples from cells treated in various ways after 60 min. exposure to $^{32}\text{PO}_4^{3-}$ were precipitated with antiserum to 113 kD protein. Lane 1, no treatment of cells; Lane 2, cells treated 7 min. with IFN- α . By comparison with the marker proteins labeled 200, 97.5, 69 and 46 kD (kilo daltons), the PO_4^{3-} labeled proteins in the precipitate are seen

25 to be 113 and 91 kD. Lane 3, cells treated with IFN- γ overnight (no phosphorylated proteins) and then (Lane 4) treated with IFN- α for 7 min. show heavier phosphorylation of 113, 91 and 84 kD.

FIGURE 12 is a chromatogram depicting the identification of phosphoamino acid.

30 Phosphate labeled protein of 113, 91 or 84 kD size was hydrolyzed and chromatographed to reveal newly labeled phosphotyrosine. Cells untreated with

IFN showed only phosphoserine label. (P Ser = phosphoserine; P Thr = phosphothreonine; P Tyr = phosphotyrosine.

FIGURE 13 depicts (A) the deduced amino acid sequence (SEQ ID NO:8) of and
5 (B-D) the DNA sequence (SEQ ID NO:7) encoding the murine 91 kD intracellular receptor recognition factor.

FIGURE 14 depicts (A) the deduced amino acid sequence (SEQ ID NO:10) of and
10 (B-C) the DNA sequence (SEQ ID NO:9) encoding the 13sf1 intracellular receptor recognition factor.

FIGURE 15 depicts (A) the deduced amino acid sequence (SEQ ID NO:12) of and
15 (B-C) the DNA sequence (SEQ ID NO:11) encoding the 19sf6 intracellular receptor recognition factor.

FIGURE 16. Determination of molecular weights of Stat91 and phospho Stat91 by native gel analysis.

A) Western blot analysis of fractions from affinity purification. Extracts from human FS2 fibroblasts treated with IFN- γ (Ext), the unbound fraction (Flow), the
20 fraction washed with Buffer AO.2 (AO.2), and the bound fraction eluted with buffer AO.8(AO.8) were immunoblotted with anti-91T.

B) Native gel analysis. Phosphorylated Stat91 (the AO.8 fraction from A) and unphosphorylated Stat91 (the Flow fraction from A) were analyzed on 4.5%, 5.5%, 6.5% and 7.5% native polyacrylamide gels followed by immunoblotting
25 with anti-91T. The top of gels (TOP) and the migration position of bromophenol blue (BPB) are indicated.

C) Ferguson plots. The relative mobilities (Rm) of the Stat91 and phospho Stat91 were obtained from Figure 1B (see Experimental Procedures). Closed circle: Chicken egg albumin (45kD); Cross: Bovine serum albumin, monomer (66 kD);
30 Open square: Bovine serum albumin, dimer (132 kD); Open circle: Urease, trimer

(272 kD); Open triangle: Unphosphorylated Stat91; Closed triangle: Phosphorylated Stat91.

- D) Determination of molecular weights from the standard curve. The molecular weights of phosphorylated and unphosphorylated Stat91 proteins (indicated as closed and open arrows, respectively) were obtained by extrapolation of their retardation coefficients.

FIGURE 17. Determination of molecular weights by glycerol gradients.

- A) Western blot analysis. Extracts from human Bud8 fibroblasts treated with IFN- γ (the rightmost lane) and every other fraction from fraction 16 to 34 were analyzed on 7.5% SDS-PAGE followed by immunoblotting with anti-91T. The peak of phosphorylated Stat91 (fraction 20) and the peak of unphosphorylated Stat91 (fraction 30) were indicated by a closed and open arrow, respectively.
- B) Mobility shift analysis. Every other fractions from the gradients were analyzed.
- C) Graphic representation of the data from A and B. Peak fraction numbers of protein standards are plotted versus their molecular weight. The position of peaks (of phosphorylated and unphosphorylated Stat91 protein are indicated by the closed and open arrows, respectively. Standards are ferritin (Fer, 440 kD), catalase (Cat, 232 kD), ferritin half unit (Fer 1/2, 220 kD), aldolase (Ald, 158 kD), bovine serum albumin (BSA, 68 kD).

FIGURE 18. Stat91 in cell extracts binds DNA as a dimer.

- A) Western blot analysis. Extracts from stable cell lines expressing either Stat84 (C84), or Stat91L (C91L) or both (Cmx) were analyzed on 7.5% SDS-PAGE followed by immunoblotting with anti-91.
- B) Gel mobility shift analysis. Extracts from stable cell lines (Fig 3A) untreated (-) or treated with IFN- γ (+) were analyzed. The positions of Stat91 homodimer (91L), Stat84 homodimer (84), and the heterodimer (84*91) are indicated.

FIGURE 19. Formation of herterodimer by denaturation and renaturation.

Cytoplasmic (Left Panel) or nuclear extracts (Right Panel) from IFN- γ -treated cell lines expressing either Stat84 (C84) or Stat91 (C91) were analyzed by gel mobility shift assays. +: with addition; -: without addition; D/R: samples were subjected to guanidinium hydrochloride denaturation and renaturation treatment.

FIGURE 20. Diagrammatic representation of dissociation and reassociation analysis.

- 10 FIGURE 21. Dissociation-reassociation analysis with peptides. Gel mobility shift analysis with IFN- γ treated nuclear extracts from cell lines expressing Stat91L (C91L, lane 15) or Stat84 (C84, lane 14) or mixture of both (lane 1-13, 16-18) in the presence of increasing concentrations of various peptides. 91-Y, unphosphorylated peptide from Stat91 (LDGPKGTGYIKTELI) (SEQ. ID NO.:18); 91Y-p, phosphotyrosyl peptide from Stat91 (GY*IKTE) (SEQ ID NO.:19); 113Y-p, phosphotyrosyl peptide with high binding affinity to Src SH2 domain (EPQY*EEIPIYL, Songyang et al., 1993, Cell 72:767-778) (SEQ. ID NO.:21). Final concentrations of peptides added: 1 μ M (lane 8), 4 μ M (lane 2,5, 11), 10 μ M (lane 9), 40 μ M (lane 3, 6, 10, 12, 14-18), 160 μ M (lane 4, 7, 13).
- 15 +: with addition; -: without addition. Right panel: antiserum tests for identity of gel-shift bands (see Figure 3).
- 20

FIGURE 22. Dissociation-reassociation analysis with GST fusion proteins. A) -

SDS-PAGE (12%) analysis of purified GST fusion proteins as visualized by

25 Commasie blue. GST-91 SH3, native SH2 domain of Stat91; GST-91 mSH2, R⁶⁰² to L⁶⁰² mutant; GST-91 SH3, SH3 domain of Stat91; GST Src SH2, the SH2 domain of src protein. Same amounts (1 μ g) of each fusion proteins were loaded. Protein markers were run in lane 1 as indicated.

B) Dissociation-reassociation analysis similar to Figure 6. Dissociating agents

- 30 were GST fusion proteins purified from bacterial expression as shown above. Final concentrations of fusion proteins added are 0.5 μ M (lanes 2, 5, 8, 11, 14),

2.5 μ M (lanes 3, 6, 9, 12, 15) and 5 μ M (lanes 4, 7, 10, 13, 17, 18). +: with addition; -: without addition; FP: fusion proteins.

FIGURE 23. Comparison of Stat91 SH2 structure with known SH2 structures.

- 5 The Stat91 sequence is disclosed herein (SEQ ID NO:4). The structures used for the other SH2s are Src (Waksman et al., 1992, Nature 358:646-653) (SEQ ID NO:22), AbI (Overduin et al., 1992, Proc. Natl. Acad. Sci. USA 89:11673-77 and 1992, Cell 70:697-704) (SEQ ID NO:23, Lck (Eck et al., 1993, Nature 362:87-91) (SEQ ID NO:24), and p85 α N (Booker et al., 1992, Nature 358:684-687) (SEQ ID NO:25). The alignment of the determined structures is by direct coordinate superimposition of the backbone structures. The names of secondary structural features and significant residues is based on the scheme of Eck et al., 1993. The boundaries and extents of the structure features are indicated by [---]. The starting numbers for the parent sequences are shown in parentheses.
- 15 Experimentally determined structurally conserved regions are from Src, p85 α , and AbI (Cowburn, unpublished). The root mean square deviation of three-dimensionally aligned structures differs by less than 1 Angstrom for the backbone non-hydrogen atoms in the sections marked by the XXX.

20

DETAILED DESCRIPTION

- In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g.,
- 25 Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual" (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" [B.D. Hames & S.J. Higgins eds. (1985)]; "Transcription And Translation" [B.D. Hames & S.J. Higgins, eds. (1984)]; "Animal Cell Culture"
 - 30 [R.I. Freshney, ed. (1986)]; "Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

The terms "receptor recognition factor", "receptor recognition-tyrosine kinase factor", "receptor recognition factor/tyrosine kinase substrate", "receptor recognition/transcription factor", "recognition factor" and "recognition factor protein(s)" and any variants not specifically listed, may be used herein interchangeably, and as used throughout the present application and claims refer to proteinaceous material including single or multiple proteins, and extends to those proteins having the amino acid sequence data described herein and presented in FIGURE 1 (SEQ ID NO:2), FIGURE 2 (SEQ ID NO:4) and in FIGURE 3 (SEQ ID NO:6), and the profile of activities set forth herein and in the Claims. Accordingly, proteins displaying substantially equivalent or altered activity are likewise contemplated. These modifications may be deliberate, for example, such as modifications obtained through site-directed mutagenesis, or may be accidental, such as those obtained through mutations in hosts that are producers of the complex or its named subunits. Also, the terms "receptor recognition factor", "recognition factor" and "recognition factor protein(s)" are intended to include within their scope proteins specifically recited herein as well as all substantially homologous analogs and allelic variations.

The amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired functional property of immunoglobulin-binding is retained by the polypeptide. NH₂ refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J. Biol. Chem.*, 243:3552-59 (1969), abbreviations for amino acid residues are shown in the following Table of Correspondence:

TABLE OF CORRESPONDENCE

<u>SYMBOL</u>		<u>AMINO ACID</u>	
	<u>1-Letter</u>	<u>3-Letter</u>	
	Y	Tyr	tyrosine
5	G	Gly	glycine
	F	Phe	phenylalanine
	M	Met	methionine
	A	Ala	alanine
	S	Ser	serine
10	I	Ile	isoleucine
	L	Leu	leucine
	T	Thr	threonine
	V	Val	valine
	P	Pro	proline
15	K	Lys	lysine
	H	His	histidine
	Q	Gln	glutamine
	E	Glu	glutamic acid
	W	Trp	tryptophan
20	R	Arg	arginine
	D	Asp	aspartic acid
	N	Asn	asparagine
	C	Cys	cysteine

- 25 It should be noted that all amino-acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-terminus to carboxy-terminus. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino-acid residues. The above Table is
- 30 presented to correlate the three-letter and one-letter notations which may appear alternately herein.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

- 5 A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

- 10 A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its either single stranded form, or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing
15 the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

- 20 An "origin of replication" refers to those DNA sequences that participate in DNA synthesis.

- A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control
25 of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA
30 sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

- 5 A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or
 10 elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and
 15 "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

- An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is
 20 "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

- A "signal sequence" can be included before the coding sequence. This sequence
 25 encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell to direct the polypeptide to the cell surface or secrete the polypeptide into the media, and this signal peptide is clipped off by the host cell before the protein leaves the cell. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes.

The term "oligonucleotide", as used herein in referring to the probe of the present invention, is defined as a molecule comprised of two or more ribonucleotides, preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

5

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a
10 nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors,
15 including temperature, source of primer and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides.

20 The primers herein are selected to be "substantially" complementary to different strands of a particular target DNA sequence. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end
25 of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to hybridize therewith and thereby form the template for the synthesis of the extension product.

30

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

- 5 A cell has been "transformed" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid.
- 10 With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a
- 15 population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

- Two DNA sequences are "substantially homologous" when at least about 75 %
- 20 (preferably at least about 80 %, and most preferably at least about 90 or 95 %) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for
 - 25 that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Maniatis et al., *supra*; DNA Cloning, Vols. I & II, *supra*; Nucleic Acid Hybridization, *supra*.

- A "heterologous" region of the DNA construct is an identifiable segment of DNA
- 30 within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian

gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains
5 introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

An "antibody" is any immunoglobulin, including antibodies and fragments thereof,
10 that binds a specific epitope. The term encompasses polyclonal, monoclonal, and chimeric antibodies, the last mentioned described in further detail in U.S. Patent Nos. 4,816,397 and 4,816,567.

An "antibody combining site" is that structural portion of an antibody molecule
15 comprised of heavy and light chain variable and hypervariable regions that specifically binds antigen.

The phrase "antibody molecule" in its various grammatical forms as used herein contemplates both an intact immunoglobulin molecule and an immunologically
20 active portion of an immunoglobulin molecule.

Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and those portions of an immunoglobulin molecule that contains the paratope, including those portions known in the art as
25 Fab, Fab', F(ab')₂ and F(v), which portions are preferred for use in the therapeutic methods described herein.

Fab and F(ab')₂ portions of antibody molecules are prepared by the proteolytic reaction of papain and pepsin, respectively, on substantially intact antibody
30 molecules by methods that are well-known. See for example, U.S. Patent No. 4,342,566 to Theofilopolous et al. Fab' antibody molecule portions are also well-

known and are produced from F(ab')₂ portions followed by reduction of the disulfide bonds linking the two heavy chain portions as with mercaptoethanol, and followed by alkylation of the resulting protein mercaptan with a reagent such as iodoacetamide. An antibody containing intact antibody molecules is preferred
 5 herein.

The phrase "monoclonal antibody" in its various grammatical forms refers to an antibody having only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically
 10 displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen; e.g., a bispecific (chimeric) monoclonal antibody.

15 The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that are physiologically tolerable and do not typically produce an allergic or similar untoward reaction, such as gastric upset, dizziness and the like, when administered to a human.

20 The phrase "therapeutically effective amount" is used herein to mean an amount sufficient to prevent, and preferably reduce by at least about 30 percent, more preferably by at least 50 percent, most preferably by at least 90 percent, a clinically significant change in the S phase activity of a target cellular mass, or other feature of pathology such as for example, elevated blood pressure, fever or
 25 white cell count as may attend its presence and activity.

A DNA sequence is "operatively linked" to an expression control sequence when the expression control sequence controls and regulates the transcription and translation of that DNA sequence. The term "operatively linked" includes having
 30 an appropriate start signal (e.g., ATG) in front of the DNA sequence to be expressed and maintaining the correct reading frame to permit expression of the

DNA sequence under the control of the expression control sequence and production of the desired product encoded by the DNA sequence. If a gene that one desires to insert into a recombinant DNA molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene.

5

The term "standard hybridization conditions" refers to salt and temperature conditions substantially equivalent to 5 x SSC and 65°C for both hybridization and wash.

- 10 In its primary aspect, the present invention concerns the identification of a receptor recognition factor, and the isolation and sequencing of a particular receptor recognition factor protein, that is believed to be present in cytoplasm and that serves as a signal transducer between a particular cellular receptor having bound thereto an equally specific polypeptide ligand, and the comparably specific
- 15 transcription factor that enters the nucleus of the cell and interacts with a specific DNA binding site for the activation of the gene to promote the predetermined response to the particular polypeptide stimulus. The present disclosure confirms that specific and individual receptor recognition factors exist that correspond to known stimuli such as tumor necrosis factor, nerve growth factor, platelet-derived
- 20 growth factor and the like. Specific evidence of this is set forth herein with respect to the interferons α and γ (IFN α and IFN γ).

- A further property of the receptor recognition factors (also termed herein signal transducers and activators of transcription -- STAT) is dimerization to form
- 25 homodimers or heterodimers upon activation by phosphorylation of tyrosine. In a specific embodiment, *infra*, Stat91 and Stat84 form homodimers and a Stat91-Stat84 heterodimer. Accordingly, the present invention is directed to such dimers, which can form spontaneously by phosphorylation of the STAT protein, or which can be prepared synthetically by chemically cross-linking two like or unlike STAT
- 30 proteins.

The present receptor recognition factor is likewise noteworthy in that it appears not to be demonstrably affected by fluctuations in second messenger activity and concentration. The receptor recognition factor proteins appear to act as a substrate for tyrosine kinase domains, however do not appear to interact with G-proteins, and therefore do not appear to be second messengers.

A particular receptor recognition factor identified herein by SEQ ID NO:4, has been determined to be present in cytoplasm and serves as a signal transducer and a specific transcription factor in response to IFN- γ stimulation that enters the nucleus of the cell and interacts directly with a specific DNA binding site for the activation of the gene to promote the predetermined response to the particular polypeptide stimulus. This particular factor also acts as a translation protein and, in particular, as a DNA binding protein in response to interferon- γ stimulation. This factor is likewise noteworthy in that it has the following characteristics:

- a) It interacts with an interferon- γ -bound receptor kinase complex;
- b) It is a tyrosine kinase substrate; and
- c) When phosphorylated, it serves as a DNA binding protein.

More particularly, the factor of SEQ ID NO:4 directly interacts with DNA after acquiring phosphate on tyrosine located at position 701 of the amino acid sequence. Also, interferon- γ -dependent activation of this factor occurs without new protein synthesis and appears within minutes of interferon- γ treatment, achieves maximum extent between 15 and 30 minutes thereafter, and then disappears after 2-3 hours.

In a particular embodiment, the present invention relates to all members of the herein disclosed family of receptor recognition factors except the 91 kD protein factors, specifically the proteins whose sequences are represented by one or more of SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

Subsequent to the filing of the initial applications directed to the present invention, the inventors have termed each member of the family of receptor recognition factors as a signal transducer and activator of transcription (STAT) protein. Each STAT protein is designated by the apparent molecular weight (*e.g.*, Stat113, Stat91, Stat84, etc.), or by the order in which it has been identified (*e.g.*, Stat1 α [Stat91], Stat1 β [Stat84], Stat2 [Stat113], Stat3 [a murine protein described in U.S. Application Serial No. 08/126,588, filed September 24, 1993 as 19sf6], and Stat4 [a murine STAT protein described in U.S. Application Serial No. 08/126,588, filed September 24, 1993 as 13sf1]). As will be readily appreciated by one of ordinary skill in the art, the choice of name has no effect on the intrinsic characteristics of the factors described herein, which were first disclosed in U.S. Application Serial No. 07/845,296, filed March 19, 1992. The present inventors have chosen to adopt this newly derived terminology herein as a convenience to the skilled artisan who is familiar with the subsequently published papers relating to the same, and in accordance with the proposal to harmonize the terminology for the novel class of proteins, and nucleic acids encoding the proteins, disclosed by the instant inventors. The terms [molecular weight] kd receptor recognition factor, Stat[molecular weight], and Stat[number] are used herein interchangeably, and have the meanings given above. For example, the terms 91 kd protein, Stat91, and Stat1 α refer to the same protein, and in the appropriate context refer to the nucleic acid molecule encoding such protein.

As stated above, the present invention also relates to a recombinant DNA molecule or cloned gene, or a degenerate variant thereof, which encodes a receptor recognition factor, or a fragment thereof, that possesses a molecular weight of about 113 kD and an amino acid sequence set forth in FIGURE 1 (SEQ ID NO:2); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 113 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA sequence shown in FIGURE 1 (SEQ ID NO:1). In another embodiment, the receptor recognition factor has a molecular weight of about 91 kD and the amino acid sequence set forth in FIGURE 2 (SEQ

ID NO:4) or FIGURE 13 (SEQ ID NO:8); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 91 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqeence shown in FIGURE 2 (SEQ ID NO:3) or FIGURE 13 (SEQ ID NO:8). In yet a further embodiment, the receptor recognition factor has a molecular weight of about 84 kD and the amino acid sequence set forth in FIGURE 3 (SEQ ID NO:6); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the 84 kD receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqeence shown in FIGURE 3 (SEQ ID NO:5). In yet another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIGURE 14 (SEQ ID NO:10); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqeence shown in FIGURE 14 (SEQ ID NO:9). In still another embodiment, the receptor recognition factor has an amino acid sequence set forth in FIGURE 15 (SEQ ID NO:12); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding such receptor recognition factor has a nucleotide sequence or is complementary to a DNA seqeence shown in FIGURE 15 (SEQ ID NO:11).

The possibilities both diagnostic and therapeutic that are raised by the existence of the receptor recognition factor or factors, derive from the fact that the factors appear to participate in direct and causal protein-protein interaction between the receptor that is occupied by its ligand, and those factors that thereafter directly interface with the gene and effect transcription and accordingly gene activation. As suggested earlier and elaborated further on herein, the present invention contemplates pharmaceutical intervention in the cascade of reactions in which the receptor recognition factor is implicated, to modulate the activity initiated by the stimulus bound to the cellular receptor.

Thus, in instances where it is desired to reduce or inhibit the gene activity resulting from a particular stimulus or factor, an appropriate inhibitor of the receptor recognition factor could be introduced to block the interaction of the receptor recognition factor with those factors causally connected with gene

- 5 activation. Correspondingly, instances where insufficient gene activation is taking place could be remedied by the introduction of additional quantities of the receptor recognition factor or its chemical or pharmaceutical cognates, analogs, fragments and the like.

- 10 As discussed earlier, the recognition factors or their binding partners or other ligands or agents exhibiting either mimicry or antagonism to the recognition factors or control over their production, may be prepared in pharmaceutical compositions, with a suitable carrier and at a strength effective for administration by various means to a patient experiencing an adverse medical condition associated
- 15 specific transcriptional stimulation for the treatment thereof. A variety of administrative techniques may be utilized, among them parenteral techniques such as subcutaneous, intravenous and intraperitoneal injections, catheterizations and the like. Average quantities of the recognition factors or their subunits may vary and in particular should be based upon the recommendations and prescription of a
- 20 qualified physician or veterinarian.

- Also, antibodies including both polyclonal and monoclonal antibodies, and drugs that modulate the production or activity of the recognition factors and/or their subunits may possess certain diagnostic applications and may for example, be
- 25 utilized for the purpose of detecting and/or measuring conditions such as viral infection or the like. For example, the recognition factor or its subunits may be used to produce both polyclonal and monoclonal antibodies to themselves in a variety of cellular media, by known techniques such as the hybridoma technique utilizing, for example, fused mouse spleen lymphocytes and myeloma cells.
- 30 Likewise, small molecules that mimic or antagonize the activity(ies) of the

receptor recognition factors of the invention may be discovered or synthesized, and may be used in diagnostic and/or therapeutic protocols.

The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal, antibody-producing cell lines can also be created by techniques other than fusion, such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., "Hybridoma Techniques" (1980); Hammerling et al., "Monoclonal Antibodies And T-cell Hybridomas" (1981); Kennett et al., "Monoclonal Antibodies" (1980); see also U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,451,570; 4,466,917; 4,472,500; 4,491,632; 4,493,890.

Panels of monoclonal antibodies produced against recognition factor peptides can be screened for various properties; i.e., isotype, epitope, affinity, etc. Of particular interest are monoclonal antibodies that neutralize the activity of the recognition factor or its subunits. Such monoclonals can be readily identified in recognition factor activity assays. High affinity antibodies are also useful when immunoaffinity purification of native or recombinant recognition factor is possible.

Preferably, the anti-recognition factor antibody used in the diagnostic methods of this invention is an affinity purified polyclonal antibody. More preferably, the antibody is a monoclonal antibody (mAb). In addition, it is preferable for the anti-recognition factor antibody molecules used herein be in the form of Fab, Fab', F(ab')₂ or F(v) portions of whole antibody molecules.

As suggested earlier, the diagnostic method of the present invention comprises examining a cellular sample or medium by means of an assay including an effective amount of an antagonist to a receptor recognition factor/protein, such as an anti-recognition factor antibody, preferably an affinity-purified polyclonal antibody, and more preferably a mAb. In addition, it is preferable for the anti-recognition factor antibody molecules used herein be in the form of Fab, Fab',

F(ab')₂ or F(v) portions or whole antibody molecules. As previously discussed, patients capable of benefiting from this method include those suffering from cancer, a pre-cancerous lesion, a viral infection or other like pathological derangement. Methods for isolating the recognition factor and inducing anti-
5 recognition factor antibodies and for determining and optimizing the ability of anti-recognition factor antibodies to assist in the examination of the target cells are all well-known in the art.

Methods for producing polyclonal anti-polypeptide antibodies are well-known in
10 the art. See U.S. Patent No. 4,493,795 to Nestor et al. A monoclonal antibody, typically containing Fab and/or F(ab')₂ portions of useful antibody molecules, can be prepared using the hybridoma technology described in *Antibodies - A Laboratory Manual*, Harlow and Lane, eds., Cold Spring Harbor Laboratory, New York (1988), which is incorporated herein by reference. Briefly, to form the
15 hybridoma from which the monoclonal antibody composition is produced, a myeloma or other self-perpetuating cell line is fused with lymphocytes obtained from the spleen of a mammal hyperimmunized with a recognition factor-binding portion thereof, or recognition factor, or an origin-specific DNA-binding portion thereof.

20 Splenocytes are typically fused with myeloma cells using polyethylene glycol (PEG) 6000. Fused hybrids are selected by their sensitivity to HAT. Hybridomas producing a monoclonal antibody useful in practicing this invention are identified by their ability to immunoreact with the present recognition factor and their ability
25 to inhibit specified transcriptional activity in target cells.

A monoclonal antibody useful in practicing the present invention can be produced by initiating a monoclonal hybridoma culture comprising a nutrient medium containing a hybridoma that secretes antibody molecules of the appropriate antigen
30 specificity. The culture is maintained under conditions and for a time period sufficient for the hybridoma to secrete the antibody molecules into the medium.

The antibody-containing medium is then collected. The antibody molecules can then be further isolated by well-known techniques.

- Media useful for the preparation of these compositions are both well-known in the art and commercially available and include synthetic culture media, inbred mice and the like. An exemplary synthetic medium is Dulbecco's minimal essential medium (DMEM; Dulbecco et al., *Virology* 8:396 (1959)) supplemented with 4.5 gm/l glucose, 20 mm glutamine, and 20% fetal calf serum. An exemplary inbred mouse strain is the Balb/c.
- Methods for producing monoclonal anti-recognition factor antibodies are also well-known in the art. See Niman et al., *Proc. Natl. Acad. Sci. USA*, 80:4949-4953 (1983). Typically, the present recognition factor or a peptide analog is used either alone or conjugated to an immunogenic carrier, as the immunogen in the before described procedure for producing anti-recognition factor monoclonal antibodies. The hybridomas are screened for the ability to produce an antibody that immunoreacts with the recognition factor peptide analog and the present recognition factor.
- The present invention further contemplates therapeutic compositions useful in practicing the therapeutic methods of this invention. A subject therapeutic composition includes, in admixture, a pharmaceutically acceptable excipient (carrier) and one or more of a receptor recognition factor, polypeptide analog thereof or fragment thereof, as described herein as an active ingredient. In a preferred embodiment, the composition comprises an antigen capable of modulating the specific binding of the present recognition factor within a target cell.

The preparation of therapeutic compositions which contain polypeptides, analogs or active fragments as active ingredients is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or

suspensions, however, solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

The active therapeutic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable
 5 excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents which enhance the effectiveness of the active ingredient.

- 10 A polypeptide, analog or active fragment can be formulated into the therapeutic composition as neutralized pharmaceutically acceptable salt forms.

Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide or antibody molecule) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or

- 15 such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

20

The therapeutic polypeptide-, analog- or active fragment-containing compositions are conventionally administered intravenously, as by injection of a unit dose, for example. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as

- 25 unitary dosage for humans, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; i.e., carrier, or vehicle.

- The compositions are administered in a manner compatible with the dosage
 30 formulation, and in a therapeutically effective amount. The quantity to be administered depends on the subject to be treated, capacity of the subject's

immune system to utilize the active ingredient, and degree of inhibition or neutralization of recognition factor binding capacity desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual. However, suitable dosages may

5 range from about 0.1 to 20, preferably about 0.5 to about 10, and more preferably one to several, milligrams of active ingredient per kilogram body weight of individual per day and depend on the route of administration. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed by repeated doses at one or more hour intervals by

10 a subsequent injection or other administration. Alternatively, continuous intravenous infusion sufficient to maintain concentrations of ten nanomolar to ten micromolar in the blood are contemplated.

The therapeutic compositions may further include an effective amount of the

15 factor/factor synthesis promoter antagonist or analog thereof, and one or more of the following active ingredients: an antibiotic, a steroid. Exemplary formulations are given below:

FormulationsIntravenous Formulation I

<u>Ingredient</u>	<u>mg/ml</u>
5 cefotaxime	250.0
receptor recognition factor	10.0
dextrose USP	45.0
sodium bisulfite USP	3.2
edetate disodium USP	0.1
10 water for injection q.s.a.d.	1.0ml

Intravenous Formulation II

<u>Ingredient</u>	<u>mg/ml</u>
ampicillin	250.0
15 receptor recognition factor	10.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0ml

20 Intravenous Formulation III

<u>Ingredient</u>	<u>mg/ml</u>
gentamicin (charged as sulfate)	40.0
receptor recognition factor	10.0
sodium bisulfite USP	3.2
25 disodium edetate USP	0.1
water for injection q.s.a.d.	1.0ml

Intravenous Formulation IV

<u>Ingredient</u>	<u>mg/ml</u>
30 recognition factor	10.0
dextrose USP	45.0

sodium bisulfite USP	3.2
edetate disodium USP	0.1
water for injection q.s.a.d.	1.0 ml

5 Intravenous Formulation V

<u>Ingredient</u>	<u>mg/ml</u>
recognition factor antagonist	5.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
10 water for injection q.s.a.d.	1.0 ml

As used herein, "pg" means picogram, "ng" means nanogram, "ug" or " μ g" mean microgram, "mg" means milligram, "ul" or " μ l" mean microliter, "ml" means milliliter, "l" means liter.

15

Another feature of this invention is the expression of the DNA sequences disclosed herein. As is well known in the art, DNA sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host.

20

Such operative linking of a DNA sequence of this invention to an expression control sequence, of course, includes, if not already part of the DNA sequence, the provision of an initiation codon, ATG, in the correct reading frame upstream of the DNA sequence.

25

A wide variety of host/expression vector combinations may be employed in expressing the DNA sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and Synthetic DNA sequences. Suitable vectors include derivatives of SV40 and known bacterial plasmids, e.g., *E. coli* plasmids col El, pCR1, pBR322, pMB9

30

and their derivatives, plasmids such as RP4; phage DNAs, e.g., the numerous derivatives of phage λ , e.g., NM989, and other phage DNA, e.g., M13 and Filamentous single stranded phage DNA; yeast plasmids such as the 2μ plasmid or derivatives thereof; vectors useful in eukaryotic cells, such as vectors useful in
 5 insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or other expression control sequences; and the like.

Any of a wide variety of expression control sequences -- sequences that control the
 10 expression of a DNA sequence operatively linked to it -- may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include, for example, the early or late promoters of SV40, CMV, vaccinia, polyoma or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the *LTR* system, the major operator and promoter regions
 15 of phage λ , the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase (e.g., Pho5), the promoters of the yeast α -mating factors, and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

20 A wide variety of unicellular host cells are also useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, fungi such as yeasts, and animal cells, such as CHO, RI.1, B-W and
 25 L-M cells, African Green Monkey kidney cells (e.g., COS 1, COS 7, BSC1, BSC40, and BMT10), insect cells (e.g., Sf9), and human cells and plant cells in tissue culture.

It will be understood that not all vectors, expression control sequences and hosts
 30 will function equally well to express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system.

However, one skilled in the art will be able to select the proper vectors, expression control sequences, and hosts without undue experimentation to accomplish the desired expression without departing from the scope of this invention. For example, in selecting a vector, the host must be considered

5 because the vector must function in it. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, will also be considered.

In selecting an expression control sequence, a variety of factors will normally be

10 considered. These include, for example, the relative strength of the system, its controllability, and its compatibility with the particular DNA sequence or gene to be expressed, particularly as regards potential secondary structures. Suitable unicellular hosts will be selected by consideration of, e.g., their compatibility with the chosen vector, their secretion characteristics, their ability to fold proteins

15 correctly, and their fermentation requirements, as well as the toxicity to the host of the product encoded by the DNA sequences to be expressed, and the ease of purification of the expression products.

Considering these and other factors a person skilled in the art will be able to

20 construct a variety of vector/expression control sequence/host combinations that will express the DNA sequences of this invention on fermentation or in large scale animal culture.

It is further intended that receptor recognition factor analogs may be prepared

25 from nucleotide sequences of the protein complex/subunit derived within the scope of the present invention. Analogs, such as fragments, may be produced, for example, by pepsin digestion of receptor recognition factor material. Other analogs, such as muteins, can be produced by standard site-directed mutagenesis of receptor recognition factor coding sequences. Analogs exhibiting "receptor

30 recognition factor activity" such as small molecules, whether functioning as promoters or inhibitors, may be identified by known *in vivo* and/or *in vitro* assays.

As mentioned above, a DNA sequence encoding receptor recognition factor can be prepared synthetically rather than cloned. The DNA sequence can be designed with the appropriate codons for the receptor recognition factor amino acid sequence. In general, one will select preferred codons for the intended host if the sequence will be used for expression. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature*, 292:756 (1981); Nambair et al., *Science*, 223:1299 (1984); Jay et al., *J. Biol. Chem.*, 259:6311 (1984).

10 Synthetic DNA sequences allow convenient construction of genes which will express receptor recognition factor analogs or "muteins". Alternatively, DNA encoding muteins can be made by site-directed mutagenesis of native receptor recognition factor genes or cDNAs, and muteins can be made directly using conventional polypeptide synthesis.

15 A general method for site-specific incorporation of unnatural amino acids into proteins is described in Christopher J. Noren, Spencer J. Anthony-Cahill, Michael C. Griffith, Peter G. Schultz, *Science*, 244:182-188 (April 1989). This method may be used to create analogs with unnatural amino acids.

20 The present invention extends to the preparation of antisense nucleotides and ribozymes that may be used to interfere with the expression of the receptor recognition proteins at the translational level. This approach utilizes antisense nucleic acid and ribozymes to block translation of a specific mRNA, either by
25 masking that mRNA with an antisense nucleic acid or cleaving it with a ribozyme.

Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule. (See Weintraub, 1990; Marcus-Sekura, 1988.) In the cell, they hybridize to that mRNA, forming a double
30 stranded molecule. The cell does not translate an mRNA in this double-stranded form. Therefore, antisense nucleic acids interfere with the expression of mRNA

into protein. Oligomers of about fifteen nucleotides and molecules that hybridize to the AUG initiation codon will be particularly efficient, since they are easy to synthesize and are likely to pose fewer problems than larger molecules when introducing them into receptor recognition factor-producing cells. Antisense methods have been used to inhibit the expression of many genes *in vitro* (Marcus-Sekura, 1988; Hambor et al., 1988).

Ribozymes are RNA molecules possessing the ability to specifically cleave other single stranded RNA molecules in a manner somewhat analogous to DNA restriction endonucleases. Ribozymes were discovered from the observation that certain mRNAs have the ability to excise their own introns. By modifying the nucleotide sequence of these RNAs, researchers have been able to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, 1988). Because they are sequence-specific, only mRNAs with particular sequences are inactivated.

Investigators have identified two types of ribozymes, *Tetrahymena*-type and "hammerhead"-type. (Hasselhoff and Gerlach, 1988) *Tetrahymena*-type ribozymes recognize four-base sequences, while "hammerhead"-type recognize eleven- to eighteen-base sequences. The longer the recognition sequence, the more likely it is to occur exclusively in the target mRNA species. Therefore, hammerhead-type ribozymes are preferable to *Tetrahymena*-type ribozymes for inactivating a specific mRNA species, and eighteen base recognition sequences are preferable to shorter recognition sequences.

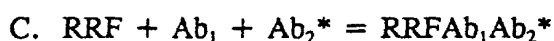
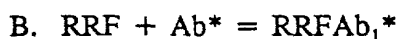
The DNA sequences described herein may thus be used to prepare antisense molecules against, and ribozymes that cleave mRNAs for receptor recognition factor proteins and their ligands.

The present invention also relates to a variety of diagnostic applications, including methods for detecting the presence of stimuli such as the earlier referenced

polypeptide ligands, by reference to their ability to elicit the activities which are mediated by the present receptor recognition factor. As mentioned earlier, the receptor recognition factor can be used to produce antibodies to itself by a variety of known techniques, and such antibodies could then be isolated and utilized as in
 5 tests for the presence of particular transcriptional activity in suspect target cells.

As described in detail above, antibody(ies) to the receptor recognition factor can be produced and isolated by standard methods including the well known hybridoma techniques. For convenience, the antibody(ies) to the receptor recognition factor
 10 will be referred to herein as Ab₁ and antibody(ies) raised in another species as Ab₂.

The presence of receptor recognition factor in cells can be ascertained by the usual immunological procedures applicable to such determinations. A number of useful
 15 procedures are known. Three such procedures which are especially useful utilize either the receptor recognition factor labeled with a detectable label, antibody Ab₁ labeled with a detectable label, or antibody Ab₂ labeled with a detectable label. The procedures may be summarized by the following equations wherein the asterisk indicates that the particle is labeled, and "RRF" stands for the receptor
 20 recognition factor:



25 The procedures and their application are all familiar to those skilled in the art and accordingly may be utilized within the scope of the present invention. The "competitive" procedure, Procedure A, is described in U.S. Patent Nos. 3,654,090 and 3,850,752. Procedure C, the "sandwich" procedure, is described in U.S. Patent Nos. RE 31,006 and 4,016,043. Still other procedures are known such as
 30 the "double antibody", or "DASP" procedure.

In each instance, the receptor recognition factor forms complexes with one or more antibody(ies) or binding partners and one member of the complex is labeled with a detectable label. The fact that a complex has formed and, if desired, the amount thereof, can be determined by known methods applicable to the detection of labels.

It will be seen from the above, that a characteristic property of Ab_2 is that it will react with Ab_1 . This is because Ab_1 raised in one mammalian species has been used in another species as an antigen to raise the antibody Ab_2 . For example, Ab_2 may be raised in goats using rabbit antibodies as antigens. Ab_2 therefore would be anti-rabbit antibody raised in goats. For purposes of this description and claims, Ab_1 will be referred to as a primary or anti-receptor recognition factor antibody, and Ab_2 will be referred to as a secondary or anti- Ab_1 antibody.

The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals which fluoresce when exposed to ultraviolet light, and others.

A number of fluorescent materials are known and can be utilized as labels. These include, for example, fluorescein, rhodamine and auramine. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate.

The receptor recognition factor or its binding partner(s) can also be labeled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. The preferred isotope may be selected from 3H , ^{14}C , ^{32}P , ^{35}S , ^{36}Cl , ^{51}Cr , ^{57}Co , ^{58}Co , ^{59}Fe , ^{90}Y , ^{125}I , ^{131}I , and ^{186}Re .

Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by

reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Many enzymes which can be used in these procedures are known and can be utilized. The preferred are peroxidase, β -glucuronidase, β -D-glucosidase, β -D-galactosidase, urease, glucose oxidase plus
5 peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090; 3,850,752; and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

A particular assay system developed and utilized in accordance with the present
10 invention, is known as a receptor assay. In a receptor assay, the material to be assayed is appropriately labeled and then certain cellular test colonies are inoculated with a quantity of both the labeled and unlabeled material after which binding studies are conducted to determine the extent to which the labeled material binds to the cell receptors. In this way, differences in affinity between materials
15 can be ascertained.

Accordingly, a purified quantity of the receptor recognition factor may be radiolabeled and combined, for example, with antibodies or other inhibitors thereto, after which binding studies would be carried out. Solutions would then be
20 prepared that contain various quantities of labeled and unlabeled uncombined receptor recognition factor, and cell samples would then be inoculated and thereafter incubated. The resulting cell monolayers are then washed, solubilized and then counted in a gamma counter for a length of time sufficient to yield a standard error of $<5\%$. These data are then subjected to Scatchard analysis after
25 which observations and conclusions regarding material activity can be drawn. While the foregoing is exemplary, it illustrates the manner in which a receptor assay may be performed and utilized, in the instance where the cellular binding ability of the assayed material may serve as a distinguishing characteristic.

30 An assay useful and contemplated in accordance with the present invention is known as a "cis/trans" assay. Briefly, this assay employs two genetic constructs,

one of which is typically a plasmid that continually expresses a particular receptor of interest when transfected into an appropriate cell line, and the second of which is a plasmid that expresses a reporter such as luciferase, under the control of a receptor/ligand complex. Thus, for example, if it is desired to evaluate a
5 compound as a ligand for a particular receptor, one of the plasmids would be a construct that results in expression of the receptor in the chosen cell line, while the second plasmid would possess a promoter linked to the luciferase gene in which the response element to the particular receptor is inserted. If the compound under test is an agonist for the receptor, the ligand will complex with the receptor, and
10 the resulting complex will bind the response element and initiate transcription of the luciferase gene. The resulting chemiluminescence is then measured photometrically, and dose response curves are obtained and compared to those of known ligands. The foregoing protocol is described in detail in U.S. Patent No. 4,981,784 and PCT International Publication No. WO 88/03168, for which
15 purpose the artisan is referred.

In a further embodiment of this invention, commercial test kits suitable for use by a medical specialist may be prepared to determine the presence or absence of predetermined transcriptional activity or predetermined transcriptional activity
20 capability in suspected target cells. In accordance with the testing techniques discussed above, one class of such kits will contain at least the labeled receptor recognition factor or its binding partner, for instance an antibody specific thereto, and directions, of course, depending upon the method selected, e.g., "competitive", "sandwich", "DASP" and the like. The kits may also contain
25 peripheral reagents such as buffers, stabilizers, etc.

Accordingly, a test kit may be prepared for the demonstration of the presence or capability of cells for predetermined transcriptional activity, comprising:

- (a) a predetermined amount of at least one labeled immunochemically reactive
30 component obtained by the direct or indirect attachment of the present receptor recognition factor or a specific binding partner thereto, to a detectable label;

- (b) other reagents; and
- (c) directions for use of said kit.

More specifically, the diagnostic test kit may comprise:

- 5 (a) a known amount of the receptor recognition factor as described above (or a binding partner) generally bound to a solid phase to form an immunosorbent, or in the alternative, bound to a suitable tag, or plural such end products, etc. (or their binding partners) one of each;
- (b) if necessary, other reagents; and
- 10 (c) directions for use of said test kit.

In a further variation, the test kit may be prepared and used for the purposes stated above, which operates according to a predetermined protocol (e.g. "competitive", "sandwich", "double antibody", etc.), and comprises:

- 15 (a) a labeled component which has been obtained by coupling the receptor recognition factor to a detectable label;
- (b) one or more additional immunochemical reagents of which at least one reagent is a ligand or an immobilized ligand, which ligand is selected from the group consisting of:
 - 20 (i) a ligand capable of binding with the labeled component (a);
 - (ii) a ligand capable of binding with a binding partner of the labeled component (a);
 - (iii) a ligand capable of binding with at least one of the component(s) to be determined; and
 - 25 (iv) a ligand capable of binding with at least one of the binding partners of at least one of the component(s) to be determined; and
- (c) directions for the performance of a protocol for the detection and/or determination of one or more components of an immunochemical reaction between
- 30 the receptor recognition factor and a specific binding partner thereto.

In accordance with the above, an assay system for screening potential drugs effective to modulate the activity of the receptor recognition factor may be prepared. The receptor recognition factor may be introduced into a test system, and the prospective drug may also be introduced into the resulting cell culture, and the culture thereafter examined to observe any changes in the transcriptional activity of the cells, due either to the addition of the prospective drug alone, or due to the effect of added quantities of the known receptor recognition factor.

PRELIMINARY CONSIDERATIONS

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain investigations with particular stimuli. Particularly, the present disclosure is illustrated by the results of work on protein factors that govern transcriptional control of IFN α -stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFN γ . The following is a brief discussion of the role that IFN is believed to play in the stimulation of transcription taken from Darnell et al. *THE NEW BIOLOGIST*, 2(10), (1990).

Activation of genes by IFN α occurs within minutes of exposure of cells to this factor (Larner et al., 1984, 1986) and is strictly dependent on the IFN α binding to its receptor, a 49-kD plasma membrane polypeptide (Uze et al., 1990). However, changes in intracellular second messenger concentrations secondary to the use of phorbol esters, calcium ionophores, or cyclic nucleotide analogs neither triggers nor blocks IFN α -dependent gene activation (Larner et al., 1984; Lew et al., 1989). No other polypeptide, even IFN γ , induces the set of interferon-stimulated genes (ISGs) specifically induced by IFN α . In addition, it has been found that IFN γ -dependent transcriptional stimulation of at least one gene in HeLa cells and in fibroblasts is also strictly dependent on receptor-ligand interaction and is not activated by induced changes in second messengers (Decker et al., 1989; Lew et al., 1989). These highly specific receptor-ligand interactions, as well as the

precise transcriptional response, require the intracellular recognition of receptor occupation and the communication to the nucleus to be equally specific.

The activation of ISGs by IFN α is carried out by transcriptional factor ISGF-3, or
 5 interferon stimulated gene factor 3. This factor is activated promptly after IFN α
 treatment without protein synthesis, as is transcription itself (Larner et al., 1986;
 Levy et al., 1988; Levy et al., 1989). ISGF-3 binds to the ISRE, the interferon-
 stimulated response element, in DNA of the response genes (Reich et al., 1987;
 Levy et al., 1988), and this binding is affected by all of an extensive set of
 10 mutations that also affects the transcriptional function of the ISRE (Kessler et al.,
 1988a). Partially purified ISGF-3 containing no other DNA-binding components
 can stimulate ISRE-dependent *in vitro* transcription (Fu et al., 1990). IFN-
 dependent stimulation of ISGs occurs in a cycle, reaching a peak of 2 hours and
 declining promptly thereafter (Larner et al., 1986). ISGF-3 follows the same
 15 cycle (Levy et al., 1988, 1989). Finally, the presence or absence of ISGF3 in a
 variety of IFN-sensitive and IFN-resistant cells correlates with the transcription of
 ISGs in these cells (Kessler et al., 1988b).

ISGF-3 is composed of two subfractions, ISGF-3 α and ISGF-3 γ , that are found in
 20 the cytoplasm before IFN binds to its receptor (Levy et al., 1989). When cells are
 treated with IFN α , ISGF-3 can be detected in the cytoplasm within a minute, that
 is, some 3 to 4 minutes before any ISGF-3 is found in the nucleus (Levy et al.,
 1989). The cytoplasmic component ISGF-3 γ can be increased in HeLa cells by
 pretreatment with IFN γ , but IFN γ does not by itself activate transcription of ISGs
 25 nor raise the concentration of the complete factor, ISGF-3 (Levy et al., 1990).
 The cytoplasmic localization of the proteins that interact to constitute ISGF-3 was
 proved by two kinds of experiments. When cytoplasm of IFN γ -treated cells that
 lack ISGF-3 was mixed with cytoplasm of IFN α -treated cells, large amounts of
 ISGF-3 were formed (Levy et al., 1989). (It was this experiment that indicated
 30 the existence of an ISGF-3 γ component and an ISGF-3 α component of ISGF-3).

In addition, Dale et al. (1989) showed that enucleated cells could respond to IFN α by forming a DNA-binding protein that is probably the same as ISGF-3.

The ISGF-3 γ component is a 48-kD protein that specifically recognizes the ISRE
 5 (Kessler et al., 1990; Fu et al., 1990). Three other proteins, presumably
 constituting the ISGF-3 α component, were found in an ISGF-3 DNA complex (Fu
 et al., 1990). The entirety of roles of, or the relationships among these three
 proteins are not yet known, but it is clear that ISGF-3 is a multimeric protein
 complex. Since the binding of IFN α to the cell surface converts ISGF-3 α from an
 10 inactive to an active status within a minute, at least one of the proteins constituting
 ISGF-3 α must be affected promptly, perhaps by a direct interaction with the IFN α
 receptor.

The details of how the ISGF-3 γ component and the three other proteins are
 15 activated by cytoplasmic events and then enter the nucleus to bind the ISRE and
 increase transcription are not entirely known. Further studies of the individual
 proteins, for example, with antibodies, are presented herein. For example, it is
 clear that, within 10 minutes of IFN α treatment, there is more ISGF-3 in the
 nucleus than in the cytoplasm and that the complete factor has a much higher
 20 affinity for the ISRE than the 48-kD ISGF-3 γ component by itself (Kessler et al.,
 1990).

In summary, the attachment of interferon- α (IFN- α) to its specific cell surface
 receptor activates the transcription of a limited set of genes, termed ISGs for
 25 "interferon stimulated genes" [Larner et al., *PROC. NATL. ACAD. SCI. USA*, 81
 (1984); Larner et al., *J. BIOL. CHEM.*, 261 (1986); Friedman et al., *CELL*, 38
 (1984)]. The observation that agents that affect second messenger levels do not
 activate transcription of these genes, led to the proposal that protein:protein
 interactions in the cytoplasm beginning at the IFN receptor might act directly in
 30 transmitting to the nucleus the signal generated by receptor occupation [Levy et
 al., *NEW BIOLOGIST*, 2 (1991)].

To test this hypothesis, the present applicants began experiments in the nucleus at the activated genes. Initially, the ISRE and ISGF-3 were discovered [Levy et al., *GENES & DEV.*, 2 (1988)].

5 Partial purification of ISGF-3 followed by recovery of the purified proteins from a specific DNA-protein complex revealed that the complete complex was made up of four proteins [Fu et al., *PROC. NATL. ACAD. SCI. USA*, 87 (1990); Kessler et al., *GENES & DEV.*, 4 (1990)]. A 48 kD protein termed ISGF-3 γ , because pre-treatment of HeLa cells with IFN- γ increased its presence, binds DNA weakly
10 on its own [*Ibid.*; and Levy et al., *THE EMBO. J.*, 9 (1990)]. In combination with the IFN- α activated proteins, termed collectively the ISGF-3 α proteins, the ISGF-3 γ forms a complex that binds the ISRE with a 50-fold higher affinity [Kessler et al., *GENES & DEV.*, 4 (1990)]. The ISGF-3 α proteins comprise a set of polypeptides of 113, 91 and 84 kD. All of the ISGF-3 components initially
15 reside in the cell cytoplasm [Levy et al., *GENES & DEV.*, 3 (1989); Dale et al., *PROC. NATL. ACAD. SCI. USA*, 86 (1989)]. However after only about five minutes of IFN- α treatment the active complex is found in the cell nucleus, thus confirming these proteins as a possible specific link from an occupied receptor to a limited set of genes [Levy et al., *GENES & DEV.*, 3 (1989)].

20

In accordance with the present invention, specific proteins comprising receptor recognition factors have been isolated and sequenced. These proteins, their fragments, antibodies and other constructs and uses thereof, are contemplated and presented herein. To understand the mechanism of cytoplasmic activation of the
25 ISGF-3 α proteins as well as their transport to the nucleus and interaction with ISGF-3 γ , this factor has been purified in sufficient quantity to obtain peptide sequence from each protein. Degenerate deoxyoligonucleotides that would encode the peptides were constructed and used in a combination of cDNA library screening and PCR amplification of cDNA products copied from mRNA to
30 identify cDNA clones encoding each of the four proteins. What follows in the examples presented herein a description of the final protein preparations that

allowed the cloning of cDNAs encoding all the proteins, and the primary sequence of the 113 kD protein arising from a first gene, and the primary sequences of the 91 and 84 kD proteins which appear to arise from two differently processed RNA products from another gene. Antisera against portions of the 84 and 91 kD

5 proteins have also been prepared and bind specifically to the ISGF-3 DNA binding factor (detected by the electrophoretic mobility shift assay with cell extracts) indicating that these cloned proteins are indeed part of ISGF-3. The availability of the cDNA and the proteins they encode provides the necessary material to understand how the liganded IFN- α receptor causes immediate cytoplasmic

10 activation of the ISGF-3 protein complex, as well as to understand the mechanisms of action of the receptor recognition factors contemplated herein. The cloning of each of ISGF3- α proteins, and the evaluation and confirmation of the particular role played by the 91 kD protein as a messenger and DNA binding protein in response to IFN- γ activation, including the development and testing of antibodies

15 to the receptor recognition factors of the present invention, are all presented in the examples that follow below.

EXAMPLE 1

20 To purify relatively large amounts of ISGF-3, HeLa cell nuclear extracts were prepared from cells treated overnight (16-18 h) with 0.5 ng/ml of IFN- γ and 45 min. with IFN- α (500u/ml). The steps used in the large scale purification were modified slightly from those described earlier in the identification of the four ISGF-3 proteins.

25 Accordingly, nuclear extracts were made from superinduced HeLa cells [Levy et al., *THE EMBO. J.*, **9** (1990)] and chromatographed as previously described [Fu et al., *PROC. NATL. ACAD. SCI. USA*, **87** (1990)] on: phosphocellulose P-11, heparin agarose (Sigma); DNA cellulose (Boehringer Mannheim; flow through was

30 collected after the material was adjusted to 0.28M KCl and 0.5% NP-40); two successive rounds of ISRE oligo affinity column (1.8 ml column, eluted with a

linear gradient of 0.05 to 1.0M KCl); a point mutant ISRE oligonucleotide affinity column (flow through was collected after the material was adjusted to 0.28M KCl); and a final round on the ISRE oligonucleotide column (material was eluted in a linear 0.05 to 1.0M NaCl gradient adjusted to 0.05% NP-40). Column

5 fractions containing ISGF-3 were subsequently examined for purity by SDS PAGE/silver staining and pooled appropriately. The pooled fractions were concentrated by a centricon-10 (Amicon). The pools of fractions from preparations 1 and 2 were combined and run on a 10 cm wide, 1.5 mm thick

10 7.5% SDS polyacrylamide gel. The proteins were electroblotted to nitrocellulose for 12 hrs at 20 volts in 12.5% MeOH, 25mM Tris, 190 mM glycine. The membrane was stained with 0.1% Ponceau Red (in 1% acetic acid) and the bands of 113 kD, 91 kD, 84 kD, and 48 kD excised and subjected to peptide analysis after tryptic digestion [Wedrychowski et al., *J. BIOL. CHEM.*, 265 (1990); Aebersold et al., *PROC. NATL. ACAD. SCI. USA*, 84 (1987)]. The resulting

15 peptide sequences for the 91 kD and 84 kD proteins are indicated in Fig. 6. Degenerate oligonucleotides were designed based on the peptide sequences t19, t13b and t27: (Forward and Reverse complements are denoted by F and R:

20 19F AACGTIGACCAATTNAACATG (SEQ ID NO:14)
 T T GC T

 T
 25 13bR GTCGATGTTNGGGTANAG (SEQ ID NO:15)
 A A A A A

27R GTACAAITCAACCAGNGCAA (SEQ ID NO:16)
 T TG T T

30 The final ISRE oligonucleotide affinity selection yielded material with the SDS polyacrylamide gel electrophoretic pattern shown in Fig. 4 (left). This gel represented about 1.5% of the available material purified from over 200 L of appropriately treated HeLa cells. While 113, 91, 84 and 48 kD bands were

35 clearly prominent in the final purified preparation (see Fig. 4, right panel), there were also two prominent contaminants of about 118 and 70 kD and a few of other

contaminants in lower amounts. [Amino acid sequence data have shown that the contaminants of 86 kD and 70 kD are the KU antigen, a widely-distributed protein that binds DNA termini. However in the specific ISGF-3: ISRE complex there is no KU antigen and therefore it has been assigned no role in IFN-dependent
 5 transcriptional stimulation, [Wedrychowski et al., *J. BIOL. CHEM.*, **265** (1990)]].

Since the mobility of the 113, 91, 84, and 48 kD proteins could be accurately marked by comparison with the partially purified proteins characterized in previous experiments [Fu et al., *PROC. NATL. ACAD. SCI. USA*, **87** (1990)],
 10 further purification was not attempted at this stage. The total purified sample from 200 L of HeLa cells was loaded onto one gel, subjected to electrophoresis, transferred to nitrocellulose and stained with Ponceau red. The 113, 84, 91, and 48 kD protein bands were separately excised and subjected to peptide analysis as described [Aebersold et al., *PROC. NATL. ACAD. SCI. USA*, **84** (1987)].
 15 Released peptides were collected, separated by HPLC and analyzed for sequence content by automated Edman degradation analysis.

Accordingly, the use of the peptide sequence data for three of four peptides from the 91 kD protein and a single peptide derived from the 84 kD protein is described
 20 herein. The peptide sequence and the oligonucleotides constructed from them are given in the legend to Fig. 4 or 6. When oligonucleotides 19F and 13bR were used to prime synthesis from a HeLa cell cDNA library, a PCR product of 475 bp was generated. When this product was cloned and sequenced it encoded the 13a peptide internally. Oligonucleotide 27R derived from the only available 84 kD
 25 peptide sequence was used in an anchored PCR procedure amplifying a 405 bp segment of DNA. This 405 bp amplified sequence was identical to an already sequenced region of the 91 kD protein. It was then realized that the peptide t27 sequence was contained within peptide t19 and that the 91 and 84 kD proteins must be related (see Fig. 5 & 7). Oligonucleotides 19F and 13a were also used to
 30 select candidate cDNA clones from a cDNA library made from mRNA prepared after 16 hr. of IFN- γ and 45 min. of IFN- α treatment.

Of the numerous cDNA clones that hybridized these oligonucleotides and also the cloned PCR products, one cDNA clone, E4, contained the largest open reading frame flanked by inframe stop codons. Sequence of peptides t19, t13a, and t13b were contained in this 2217 bp ORF (see Fig. 6) which was sufficient to encode a protein of 739 amino acids (calculated molecular weight of 86 kD). The codon for the indicated initial methionine was preceded by three in frame stop codons. This coding capacity has been confirmed by translating in vitro an RNA copy of the E4 clone yielding product of nominal size of 86 kD, somewhat shorter than the *in vitro* purified 91 kD protein discussed earlier (data not shown). Perhaps this result indicates post-translational modification of the protein in the cell.

A second class of clones was also identified (see Fig. 5). E3, the prototype of this class was identical to E4 from the 5' end to bp 2286 (aa 701) at which point the sequences diverged completely. Both cDNAs terminated with a poly(A) tail.

Primer extension analysis suggested another ~150 bp were missing from the 5' end of both mRNAs. DNA probes were made from the clones representing both common and unique sequences for use in Northern blot analyses. The preparation of the probes is as follows: 20 mg of cytoplasmic RNA (0.5% NP-40 lysate) of IFN- α treated (6 h) HeLa RNA was fractionated in a 1% agarose, 6% formaldehyde gel (in 20 mM MOPS, 5mM NaAc, 1 mM EDTA, pH 7.0) for 4.5 h at 125 volts. The RNA was transferred in 20 x SSC to Hybond-N (Amersham), UV crosslinked and hybridized with 1x10⁶ cpm/ml of the indicated probes (1.5x10⁸ cpm/mg).

Probes from regions common to E3 and E4 hybridized to two RNA species of approximately 3.1 KB and 4.4 KB. Several probes derived from the 3' non-coding end of E4, which were unique to E4, hybridized only the larger RNA species. A labeled DNA probe from the unique 3' non-coding end of E3 hybridized only the smaller RNA species.

Review of the sequence at the site of 3' discontinuity between E3 and E4 suggested that the shorter mRNA results from choice of a different poly(A) site and 3' exon that begins at bp 2286 (the calculated molecular weight from the E3. The last two nucleotides before the change are GT followed by GT in E3 in line
 5 with the consensus nucleotides at an exon-intron junction. Since the ORF of E4 extends to bp 2401 it encodes a protein that is 38 amino acids longer than the one encoded by E3, but is otherwise identical (ORF is 82 kD).

Since there is no direct assay for the activity of the 91 or 84 kD protein, an
 10 independent method was needed to determine whether the cDNA clones we had isolated did indeed encode proteins that are part of ISGF-3. For this purpose antibodies were initially raised against the sequence from amino acid 597 to amino acid 703 (see Fig. 6) by expressing this peptide in the pGEX-3X vector (15) as a bacterial fusion protein. This antiserum (a42) specifically recognized the 91 kD
 15 and 84 kD proteins in both crude extracts and purified ISGF-3 (see Fig. 7a). More importantly this antiserum specifically affected the ISGF-3 band in a mobility shift assay using the labeled ISRE oligonucleotide (see Fig. 7b) confirming that the isolated 91 kD and 84 kD cDNA clones (E4 and E3) represent a component of ISGF-3. Additional antisera were raised against the amino
 20 terminus and carboxy terminus of the protein encoded by E4. The amino terminal 59 amino acids that are common to both proteins and the unique carboxy terminal 34 amino acids encoded only by the larger mRNA were expressed as fusion proteins in pGEX-3X for immunization of rabbits. Western blot analysis with highly purified ISGF-3 demonstrated that the amino terminal antibody (a55)
 25 recognized both the 91 kD and 84 kD proteins as expected. However, the other antibody (a57) recognized only the 91 kD protein confirming our assumption that the larger mRNA (4.4 KB) and larger cDNA encodes the 91 kD protein while the shorter mRNA (3.1 KB) and cDNA encodes the 84 kD protein (see Fig. 7a).

In this example, the cloning of the 113 kD protein that comprises one of the three ISGF-3 α components is disclosed.

From SDS gels of highly purified ISGF-3, the 113 kD band was identified,
 5 excised and subjected to cleavage and peptide sequence analysis [Aebersold et al.,
PROC. NATL. ACAD. SCI. USA, **87** (1987)]. Five peptide sequences (A-E) were
 obtained (Fig. 8A). Degenerate oligonucleotide probes were designed according to
 these peptides which then were radiolabeled to search a human cDNA library for
 clones that might encode the 113 kD protein. Eighteen positive cDNA clones
 10 were recovered from 2.5×10^5 phage plaques with the probe derived from peptide
 E (Fig. 8A, and the legend). Two of them were completely sequenced. Clone f11
 contained a 3.2 KB cDNA, and clone ka31 a 2.6 KB cDNA that overlapped about
 2 KB but which had a further extended 5' end in which a candidate AUG initiation
 codon was found associated with a well-conserved Kozak sequence [Kozak,
 15 *NUCLEIC ACIDS RES.*, **12** (1984)].

In addition to the phage cDNA clones, a PCR product made between
 oligonucleotides that encoded peptide D and E also yielded a 474 NT fragment
 that when sequenced was identical with the cDNA clone in this region. A
 20 combination of these clones f11 and ka31 revealed an open reading frame capable
 of encoding a polypeptide of 851 amino acids (Fig. 8A). These two clones were
 joined within their overlapping region and RNA transcribed from this recombinant
 clone was translated in vitro yielding a polypeptide that migrated in an SDS gel
 with a nominal molecular weight of 105 kD (Fig. 9A). An appropriate clone
 25 encoding the 91 kD protein was also transcribed and the RNA translated in the
 same experiment. Since both the apparently complete cDNA clones for the 113
 kD protein and the 91 kD protein produce RNAs that when translated into proteins
 migrate somewhat faster than the proteins purified as ISGF-3 components, it is
 possible that the proteins undergo post-translational modification in the cell causing
 30 them to be slightly retarded during electrophoresis. When a 660 bp cDNA

encoding the most 3' end of the 113 kD protein was used in a Northern analysis, a single 4.8 KB mRNA species was observed (Figure 9B).

- No independent assay is known for the activity of the 113 kD (or indeed any of the ISGF-3 α proteins,) but it is known that the protein is part of a DNA binding complex that can be detected by an electrophoretic mobility shift assay [Fu et al., *PROC. NATL. ACAD. SCI. USA*, 87 (1990)]. Antibodies to DNA binding proteins are known to affect the formation or migration of such complexes. Therefore antiserum to a polypeptide segment (amino acid residues 323 to 527) fused with bacterial glutathione synthetase [Smith et al., *PROC. NATL. ACAD. SCI. USA*, 83 (1986)] was raised in rabbits to determine the reactivity of the ISGF-3 proteins with the antibody. A Western blot analysis showed that the antiserum reacted predominantly with a 113 kD protein both in the ISGF3 fraction purified by specific DNA affinity chromatography (Lane 1) and in crude cell extract (Lane 2, Fig. 10A). The weak reactivity to lower protein bands was possibly due to 113 kD protein degradation. Most importantly, the antiserum specifically removed almost all of the gel-shift complex leaving some of the oligonucleotide probe in "shifted-shift" complexes which were specifically competed away with a 50 fold molar excess of the oligonucleotide binding site (the ISRE, ref. 2) for ISGF3 (Fig. 10B). Notably, this antiserum had no effect on the faster migrating shift band produced by ISGF3- γ component alone (Figure 10B). Thus it appeared that the antiserum to the 113 kD fusion product does indeed react with another protein that is part of the complete ISGF-3 complex.
- A detailed sequence comparison between the 113 and 91 sequences followed (Fig. 8B): while the nucleotide sequence showed only a distant relationship between the two proteins, there were long stretches of amino acid identity. These conserved regions were scattered throughout almost the entire 715 amino acid length encoded by the 91/84 clone. It was particularly striking that the regions corresponding to amino acids 1 to 48 and 317 to 353 and 654 to 678 in the 113 sequence were 60%

to 70% identical to corresponding regions of the 91 kD sequence. Thus the genes encoding the 113 and 84/91 proteins are closely related but not identical.

- Through examination for possible consensus sequences that might reveal
- 5 sub-domain structures in the 113 kD or 84/91 kD sequence, it was found that both proteins contained regions whose sequence might form a coil structure with heptad leucine repeats. This occurred between amino acid 210 and 245 in the 113 kD protein and between 209 and 237 in the 84/91 protein. In both the 113 kD and the 91/84 kD sequences, 4 out of 5 possible heptad repeats were leucine and one was
 - 10 valine. Domains of this type might provide a protein surface that encourages homo-or heterotypic protein interactions which have been observed in several other transcription factors [Vinson et al., *SCIENCE*, 246 (1989)]. An extended acidic domain was located at the carboxyl terminal of the 113 kD protein but not in 91 kD protein (Fig. 8A), possibly implicating the 113 kD protein in gene activation
 - 15 [Hope et al., Ma et al., *CELL*, 48 (1987)].

DISCUSSION

- When compared at moderate or high stringency to the Genbank and EMBL data bases, there were no sequences like 113 or the 84/91 sequence. Preliminary PCR
- 20 experiments however indicate that there are other family members with different sequences recoverable from a human cell cDNA library (Qureshi and Darnell unpublished). Thus, it appears that the 113 and 84/91 sequences may represent the first two members to be cloned of a larger family of proteins. We would hypothesize that the 113 kD and 84/91 kD proteins may act as signal transducers,
 - 25 somehow interacting with the internal domain of a liganded IFN α receptor or its associated protein and further that a family of waiting cytoplasmic proteins exist whose purpose is to be specific signal transducers when different receptors are occupied. Many experiments lie ahead before this general hypothesis can be crucially tested. Recent experiments have indicated that inhibitors of protein
 - 30 kinases can prevent ISGF-3 complex formulation [Reich et al., *PROC. NATL. ACAD. SCI. USA*, 87 (1990); Kessler et al., *J. BIOL. CHEM.*, 266 (1991)].

However, neither the IFN α or IFN γ receptors that have so far been cloned have intrinsic kinase activity [Uze et al., *CELL*, **60** (1990); Aguet et al., *CELL*, **55** (1988)]. We would speculate that either a second receptor chain with kinase activity or a separate kinase bound to a liganded receptor could be a part of a
5 complex that would convey signals to the ISGF-3 α proteins at the inner surface of the plasma membrane.

From the above, it has been concluded that accurate peptide sequence from ISGF-3 protein components have been determined, leading to correct identification
10 of cDNA clones encoding the 113, 91 and 84 kD components of ISGF-3 . Since staurosporine, a broadly effective kinase inhibitor blocks IFN- α induction of transcription and of ISGF-3 formation [Reich et al., *PROC. NATL. ACAD. SCI. USA*, **87** (1990); Kessler et al., *J. BIOL. CHEM.*, **266** (1991)] it seems possible that the ISGF-3 α proteins are direct cytoplasmic substrates of a liganded
15 receptor-associated kinase. The antiserum against these proteins should prove invaluable in identifying the state of the ISGF-3 α proteins before and after IFN treatment and will allow the direct exploration of the biochemistry of signal transduction from the IFN receptor.

20

EXAMPLE 3

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain investigations with particular stimuli. Particularly, the present disclosure is
25 illustrated by the results of work on protein factors that govern transcriptional control of IFN α -stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFN γ .

For example, there is evidence that the 91 kD protein is the tyrosine kinase target
30 when IFN γ is the ligand. Thus two different ligands acting through two different receptors both use these family members. With only a modest number of family

members and combinatorial use in response to different ligands, this family of proteins becomes an even more likely possibility to represent a general link between ligand-occupied receptors and transcriptional control of specific genes in the nucleus.

5

Further study of the 113, 91 and 84 kD proteins of the present invention has revealed that they are phosphorylated in response to treatment of cells with IFN α (Figure 11). Moreover, when the phosphoamino acid is determined in the newly phosphorylated protein the amino acid has been found to be tyrosine (Fig. 12).

10 This phosphorylation has been observed to disappear after several hours, indicating action of a phosphatase of the 113, 91 and 84 kD proteins to stop transcription. These results show that IFN-dependent transcription very likely demands this particular phosphorylation and a cycle of interferon-dependent phosphorylation-dephosphorylation is responsible for controlling transcription.

15

It is proposed that other members of the 113-91 protein family will be identified as phosphorylation targets in response to other ligands. If as is believed, the tyrosine phosphorylation site on proteins in this family is conserved, one can then easily determine which family members are activated (phosphorylated), and likewise the particular extracellular polypeptide ligand to which that family member is

20 responding. The modifications of these proteins (phosphorylation and dephosphorylation) enables the preparation and use of assays for determining the effectiveness of pharmaceuticals in potentiating or preventing intracellular responses to various polypeptides, and such assays are accordingly contemplated

25 within the scope of the present invention.

EXAMPLE 4

Identification of murine 91 kD protein

30 A fragment of the gene encoding the human 91 kD protein was used to screen a murine thymus and spleen cDNA library for homologous proteins. The screening

assay yielded a highly homologous gene encoding a murine polypeptide that is greater than 95% homologous to the human 91 kD protein. The nucleic acid and deduced amino acid sequence of the murine 91 kD protein are shown in Figure 12A-12C, and SEQ ID NO:7 (nucleotide sequence) and SEQ ID NO:8 (amino acid sequence).

EXAMPLE 5

Additional Members of The 113 - 91 Protein Family

- 10 Using a 300 nuclide fragment amplified by PCR from the SH2 region of the murine 91kD protein gene, murine genes encoding two additional members of the 113-91 family of receptor recognition factor proteins were isolated from a murine splenic/thymic cDNA library according to the method of Sambrook et al. (1989, *Molecular Cloning, A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Press: Cold Spring Harbor, New York) constructed in the ZAP vector. Hybridization was carried out at 42°C and washed at 42°C before the first exposure (Church and Gilbert, 1984, Proc. Natl. Acad. Sci. USA 81:1991-95). Then the filters were washed in 2X SSC, 0.1% SDS at 65°C for a second exposure. Stat1 clones survived the 65°C washing, whereas Stat3 and Stat4 clones were identified as
- 15 Cold Spring Harbor, New York) constructed in the ZAP vector. Hybridization was carried out at 42°C and washed at 42°C before the first exposure (Church and Gilbert, 1984, Proc. Natl. Acad. Sci. USA 81:1991-95). Then the filters were washed in 2X SSC, 0.1% SDS at 65°C for a second exposure. Stat1 clones survived the 65°C washing, whereas Stat3 and Stat4 clones were identified as
- 20 Stat1 clones survived the 65°C washing, whereas Stat3 and Stat4 clones were identified as plaques that lost signals at 65°C. The plaques were purified and subcloned according to Stratagene commercial protocols.

This probe was chosen to screen for other STAT family members because, while Stat1 and Stat2 SH2 domains are quite similar over the entire 100 to 120 amino acid region, only the amino terminal half of the STAT SH2 domains strongly resemble the SH2 regions found in other proteins.

The two genes have been cloned into plasmids 13sf1 and 19sf6. The nucleotide sequence, and deduced amino acid sequence, for the 13sf1 and 19sf6 genes are shown in Figures 14 and 15, respectively. These proteins are alternatively termed Stat4 and Stat3, respectively.

Comparison with the sequence of Stat91 (Stat1) and Stat113 (Stat2) shows several highly conserved regions, including the putative SH3 and SH2 domains. The conserved amino acid stretches likely point to conserved domains that enable these proteins to carry out transcription activation functions. Stat3, like Stat1 (Stat91), is widely expressed, while Stat4 expression is limited to the testes, thymus, and spleen. Stat3 has been found to be activated as a DNA binding protein through phosphorylation on tyrosine in cells treated with EGF or IL-6, but not after IFN- γ , treatment.

- Both the 13sf1 and 19sf6 genes share a significant homology with the genes encoding the human and murine 91 kD protein. There is corresponding homology between the deduced amino acid sequences of the 13sf1 and 19sf6 proteins and the amino acid sequences of the human and murine 91 kD proteins, although not the greater than 95% amino acid homology that is found between the murine and human 91 kD proteins. Thus, though clearly of the same family as the 91 kD protein, the 13sf1 and 19sf6 genes encode distinct proteins.

The chromosomal locations of the murine STAT proteins (1-4) have been determined: Stat1 and Stat4 are located in the centromeric region of mouse chromosome 1 (corresponding to human 2q 32-34q); the two other genes are on other chromosomes.

Southern analysis using probes derived from 13sf1 and 19sf6 on human genomic libraries have established that genes corresponding to the murine 13sf1 and 19sf6 genes are found in humans.

Tissue distribution of mRNA expression of these genes was evaluated by Northern hybridization analysis. The results of this distribution analysis are shown in the following Table.

TABLE

DISTRIBUTION OF mRNA EXPRESSION OF 13sf1, 19sf6, 91 kD PROTEINS

ORGAN	13sf1	19sf6	91 KD
BRAIN	-	+	-
5 HEART	-	+++	-
KIDNEY	-	-	-
LIVER	-	+	+
LUNG	-	-	-
SPLEEN	+	+	++++
10 TESTIS	++++	++	N.A.
THYMUS	++	++	+++
EMBRYO (16d)	not found	found	found

Northern analysis demonstrates that there is variation in the tissue distribution of expression of the mRNAs encoded by these genes. The variation and tissue distribution indicates that the specific genes encode proteins that are responsive to different factors, as would be expected in accordance with the present invention. The actual ligand, the binding of which induces phosphorylation of the newly discovered factors, will be readily determinable based on the tissue distribution evidence described above.

To determine whether the Stat3 and Stat4 proteins were present in cells, protein blots were carried out with antisera against each protein. The antisera were obtained by subcloning amino acids 688 to 727 of Stat3 and 678 to 743 of Stat4 to pGEX11t (Pharmacia) by PCR with oligonucleotides based on the boundary sequence plus restriction sites (BamHI at the 5' end and EcoRI at the 3' end), allowing for in-frame fusion with GST. One milligram of each antigen was used

for the immunization and three booster injections were given 4 weeks apart. Anti-Stat3 and anti-Stat4 sera were used 1:1000 in Western blots using standard protocols. To avoid cross reactivity of the antisera, antibodies were raised against the C-terminal of Stat3 and Stat4, the less homologous region of the protein.

5

These proteins were unambiguously found in several tissues where the mRNA was known to be present. Protein expression was checked in several cell lines as well. A protein of 89 kD reactive with Stat4 antiserum was expressed in 70Z cells, a preB cell line, but not in many other cell lines. Stat3 was highly expressed, predominantly as a 97 kD protein, in 70Z, HT2 (a mouse helper T cell clone), and U937 (a macrophage-derived cell).

10

To prove that the full length functional cDNA clones of Stat3 and Stat4 were obtained, the open reading frames of each cDNA was independently (*i.e.*, separately) cloned into the Rc/CMV expression vector (Invitrogen) downstream of a CMV promoter. The resulting plasmids were transfected into COS1 cells and proteins were extracted 60 hrs post-transfection and examined by Western blot after electrophoresis. Untransfected COS1 cells expressed a low level of 97 kD Stat3 protein but did not express a detectable level of Stat4. Upon transfection of the Stat3-expressing plasmid, the 97 kD Stat3 was increased at least 10-fold. And 89 kD protein antigenically related to Stat3, found as a minor band in most cell line extracts, was also increased post-transfection. This protein therefore appears to represent another form of Stat3 protein, or an antigenically similar protein whose synthesis is stimulated by Stat3. Transfection with Stat4 led to the expression of a 89 kD reactive band indistinguishable in size from the p89 Stat4 found in 70Z cell extracts.

15

20

25

DISCUSSION

30

As mentioned earlier, the observation and conclusion underlying the present invention were crystallized from a consideration of the results of certain

investigations with particular stimuli. Particularly, the present disclosure is illustrated by the results of work on protein factors that govern transcriptional control of IFN α -stimulated genes, as well as more recent data on the regulation of transcription of genes stimulated by IFN γ . The present disclosure is further
5 illustrated by the identification of related genes encoding protein factors responsive to as yet unknown factors. It is expected that the murine 91 kD protein is responsive to IFN- γ .

For example, the above represents evidence that the 91 kD protein is the tyrosine
10 kinase target when IFN γ is the ligand. Thus two different ligands acting through two different receptors both use these family members. With only a modest number of family members and combinatorial use in response to different ligands, this family of proteins becomes an even more likely possibility to represent a general link between ligand-occupied receptors and transcriptional control of
15 specific genes in the nucleus.

It is proposed and shown by the foregoing that other members of the 113-91 protein family will be and have been identified as phosphorylation targets in response to other ligands. If as is believed, the tyrosine phosphorylation site on
20 proteins in this family is conserved, one can then easily determine which family members are activated (phosphorylated), and likewise the particular extracellular polypeptide ligand to which that family member is responding. The modifications of these proteins (phosphorylation and dephosphorylation) enables the preparation and use of assays for determining the effectiveness of pharmaceuticals in
25 potentiating or preventing intracellular responses to various polypeptides, and such assays are accordingly contemplated within the scope of the present invention.

Earlier work has concluded that DNA binding protein was activated in the cell cytoplasm in response to IFN- γ treatment and that this protein stimulated
30 transcription of the GBP gene (10,14). In the present work, with the aid of antisera to proteins originally studied in connection with IFN- α gene stimulation

(7,12,15), the 91 kD ISGF-3 protein has been assigned a prominent role in IFN- γ gene stimulation as well. The evidence for this conclusion included: 1) antisera specific to the 91 kD protein affected the IFN- γ dependent gel-shift complex, and 2) A 91 kD protein could be cross-linked to the GAS IFN- γ activated site. 3) A 5 ^{35}S -labeled 91 kD protein and a 91 kD immunoreactive protein specifically purified with the gel-shift complex. 4) The 91 kD protein is an IFN- γ dependent tyrosine kinase substrate as indeed it had earlier proved to be in response to IFN- α (15). 5) The 91 kD protein but not the 113 kD protein moved to the nucleus in response to IFN- γ treatment. None of these experiments prove but do strongly suggest that 10 the same 91 kD protein acts differently in different DNA binding complexes that are triggered by either IFN- α or IFN- γ .

These results strongly support the hypothesis originated from studies on IFN- α that polypeptide cell surface receptors report their occupation by extracellular ligand to 15 latent cytoplasmic proteins that after activation move to the nucleus to trigger transcription (4,15,21). Furthermore, because cytoplasmic phosphorylation and factor activation is so rapid it appears likely that the functional receptor complexes contain tyrosine kinase activity. Since the IFN- γ receptor chain that has been cloned thus far (22) has no hint of possessing intrinsic kinase activity, perhaps 20 some other molecule with tyrosine kinase activity couples with the IFN- γ receptor. Two recent results with other receptors suggest possible parallels to the situation with the IFN receptors. The *trk* protein which has an intracellular tyrosine kinase domain, associates with the NGF receptor when that receptor is occupied (23). In addition, the *lck* protein, a member of the *src* family of tyrosine kinases, is 25 co-precipitated with the T cell receptor (24). It is possible to predict that signal transduction to the nucleus through these two receptors could involve latent cytoplasmic substrates that form part of activated transcription factors. In any event, it seems possible that there are kinases like *trk* or *lck* associated with the IFN- γ receptor or with IFN- α receptor.

With regard to the effect of phosphorylation on the 91 kD protein, it was something of a surprise that after IFN- γ treatment the 91 kD protein becomes a DNA binding protein. Its role must be different in response to IFN- α treatment. Tyrosine is also phosphorylated on tyrosine and joins a complex with the 113 and 84 kD proteins but as judged by UV cross-linking studies (7), the 91 kD protein does not contact DNA.

In addition to becoming a DNA binding protein it is clear that the 91 kD protein is specifically translocated the nucleus in the wake of IFN- γ stimulation.

10

EXAMPLE : DIMERIZATION OF PHOSPHORYLATED STAT91

Stat91 (a 91 kD protein that acts as a signal transducer and activator of transcription) is inactive in the cytoplasm of untreated cells but is activated by phosphorylation on tyrosine in response to a number of polypeptide ligands including IFN- α and IFN- γ . This example reports that inactive Stat91 in the cytoplasm of untreated cells is a monomer and upon IFN- γ induced phosphorylation it forms a stable homodimer. The dimer is capable of binding to a specific DNA sequence directing transcription. Dissociation and reassociation assays show that dimerization of Stat91 is mediated through SH2-phosphotyrosyl peptide interactions. Dimerization involving SH2 recognition of specific phosphotyrosyl peptides may well provide a prototype for interactions among family members of STAT proteins to form different transcription complexes and Jak2 for the IFN- γ pathway (42, 43, 44). These kinases themselves become tyrosine phosphorylated to carry out specific signaling events.

Materials and Methods

Cell Culture. Human 2fTGH, U3A cells were maintained in DMEM medium supplied with 10% bovine calf serum. U3A cell lines supplemented with various Stat91 protein constructs were maintained in 0.1 mg/ml G418 (Gibco, BRL).

Stable cell lines were selected as described (45). IFN- γ (5 ng/ml, gift from Amgen) treatment of cells was for 15 min. unless otherwise noted.

Plasmid Constructions. Expression construct MNC-84 was made by insertion of the cDNA into the Not I-Bam HI cloning site of an expression vector PMNC (45, 35). MNC-91L was made by insertion of the Stat91 cDNA into the Not I -Bam HI cloning sites of pMNC without the stop codon at the end, resulting the production of a long form of Stat91 with a C-terminal tag of 34 amino acids encoded by PMNC vector.

10

GST fusion protein expression plasmids were constructed by the using the pGEX-2T vector (Pharmacia). GST-91SH2 encodes amino acids 573 to 672 of Stat91; GST-91mSH2 encodes amino acids 573 to 672 of Stat91 with an Arg-602- \rightarrow Leu-602 mutation; and GST-91SH3 encodes amino acids 506 to 564 of Stat91.

15

DNA Transfection. DNA transfection was carried by the calcium phosphate method, and stable cell lines were selected in Dulbecco's modified Eagle's medium containing G418 (0.5 mg/ml, Gibco), as described (45).

20 *Preparation of Cell Extracts.* Crude whole cell extracts were prepared as described (31). Cytoplasmic and nuclear extracts were prepared essentially as described (46).

25 *Affinity Purification.* Affinity purification with a biotinylated oligonucleotide was described (31). The sequence of the biotinylated GAS oligonucleotide was from the Ly6E gene promoter (34).

30 *Nondenaturing Polyacrylamide Gel Analysis.* A nondenatured protein molecular weight marker kit with a range of molecular weights from 14 to 545 kD was obtained from Sigma. Determining molecular weights using nondenaturing polyacrylamide gel was carried out following the manufacturer's procedure, which

is a modification of the methods of Bryan and Davis (47, 48). Phosphorylated and unphosphorylated Stat91 samples obtained from affinity purification using a biotinylated GAS oligonucleotide (31) were resuspended in a buffer containing 10 mM Tris (pH 6.7), 16% glycerol, 0.04% bromphenol blue (BPB). The mixtures
 5 were analyzed on 4.5%, 5.5%, 6.5%, and 7.5.% native gels side by side with standard markers using a Bio-Rad mini-Protean II Cell electrophoresis system. Electrophoresis was stopped when the dye (BPB) reached the bottom of the gels. The molecular size markers were revealed by Coomassie blue staining. Phosphorylated and unphosphorylated Stat91 samples were detected by
 10 immunoblotting with anti-91T.

Glycerol Gradient Analysis. Cells extracts (Bud 8) were mixed with protein standards (Pharmacia) and subjected to centrifugation through preformed 10%-40% glycerol gradients for 40 hours at 40,000 rpm in an SW41 rotor as described
 15 (6).

Gel Mobility Shift Assays. Gel mobility shift assays were carried out as described (34). An oligonucleotide corresponding to the GAS element from the human Fc γ RI receptor gene (Pearse et al. 1993) was synthesized and used for gel
 20 mobility shift assays. The oligonucleotide has the following sequence: 5'GATCGAGATGTATTTCCAGAAAAG3' (SEQ. ID NO:17).

Synthesis of Peptides. Solid phase peptide synthesis was used with either a DuPont RAMPS multiple synthesizer or by manual synthesis. C-terminal amino
 25 attached to Wang resin were obtained from DuPont/NEN. All amino acids were coupled as the N-Fmoc pentafluorophenyl esters (Advanced Chemtech), except for N-Fmoc, PO-dimethyl-L-phosphotyrosine (Bachem). Double couplings were used. Cleavage from resin and deprotection used thioanisol/m-cresol/TFA/TMSBr at 4°C for 16 hr. Purification used C-18 column HPLC with 0.1% TFA/acetonitrile
 30 gradients. Peptides were characterized by ¹H and ³¹P NMR, and by Mass Spec, and were greater than 95% pure.

Guanidium Hydrochloride Treatment. Extracts were incubated with guanidium hydrochloride (final concentration was 0.4 to 0.6 M) for two min. at room temperature and then diluted with gel shift buffer (final concentration of guanidium hydrochloride was 100 mM) and incubated at room temperature for 15 min. ³²P-

5 labeled GAS oligonucleotide probe was then added directly to the mixture followed by gel mobility shift assay.

Dissociation-reassociation Analysis. Extracts were incubated with various concentrations of peptides or fusion proteins, and ³²P-labeled GAS oligonucleotide probe in gel shift buffer was then added to promote the formation of protein-DNA complex followed by mobility shift analysis. This assay did not involve guanidium hydrochloride treatment.

Preparation of Fusion Proteins. Bacterially expressed GST fusion proteins were purified using standard techniques, as described in Birge et al., 1992. Fusion proteins were quantified by O.D. absorbance at 280nm. Aliquotes were frozen at -70°C.

Results

20 *Detection of Ligand Induced Dimer Formation of Stat91 in Solution.* In untreated cells, Stat91 is not phosphorylated on tyrosine. Treatment with IFN-γ leads within minutes to tyrosine phosphorylation and activation of DNA binding capacity. The phosphorylated form migrates more slowly during electrophoresis under denaturing conditions affording a simple assay for the phosphoprotein (31).

To determine the native molecular weights of the phosphorylated and unphosphorylated forms of Stat91, we separated them by affinity purification using a biotinylated deoxyoligonucleotide containing a GAS sequence (interferon gamma activation site) (Figure 16A). The separation of phosphorylated Stat91 from the unphosphorylated form was efficient as almost all detectable phosphorylated form

could bind to the GAS site while unphosphorylated Stat91 remained unbound. To determine the molecular weights of the purified phosphorylated Stat91 and unphosphorylated Stat91, samples of each were then subjected to electrophoresis through a set of nondenaturing gels containing various concentrations of acrylamide followed by Western blot analysis (Figure 16B). Native protein size markers (Sigma) were included in the analysis.

This technique was originally described by Bryan (48) and was recently used for dimer analysis (49). The logic of the technique is that increasing gel concentrations affect the migration of larger proteins more than smaller proteins, and the analysis is not affected by modifications such as protein phosphorylation (49).

A function of the relative mobilities (R_m) was plotted versus the concentration of acrylamide for each sample to construct Ferguson plots (Figure 16C). The logarithm of the retardation coefficient (calculated from Figure 16C) of each sample was then plotted against the logarithm of the relevant molecular weight range (Figure 16D). By extrapolation of its retardation coefficient (Figure 16D), the native molecular weight of Stat91 from untreated cells was estimated to be approximately 95 kD, while tyrosine phosphorylated Stat91 was estimated to be about twice as large, or approximately 180 kD. Because the calculated molecular weight from amino acid sequence of Stat91 is 87 kD, and Stat91 migrates on denaturing SDA gels with an apparent molecular weight of 91 kD (see *supra*, and refs. 12 and 45), we concluded that in solution, unphosphorylated Stat91 existed as a monomer while tyrosine phosphorylated Stat91 is a dimer.

We also employed glycerol gradient analysis to estimate the native molecular weights of both phosphorylated and unphosphorylated Stat91 (Figure 17). Whole cell extract of fibroblast cells (Bud8) treated with IFN- γ were prepared and subjected to sedimentation through a 10-40% glycerol gradient. Fractions from the gradient were collected and analyzed by both immunoblotting and gel mobility

shift analysis (Figure 17A and 17B). As expected, two electrophoretic forms of Stat91 could be detected by immunoblotting (Figure 17A): the slow-migrating form (tyrosine phosphorylated) and the fast-migrating form (unphosphorylated; Figure 17A). The phosphorylated Stat91 sedimented more rapidly than the

5 unphosphorylated form. Again, using molecular weight markers, the native molecular weight of the unphosphorylated form of Stat91 appeared to be about 90 kD while the tyrosine phosphorylated form of Stat91 was about 180kD (Figure 17C), supporting the conclusion that unphosphorylated Stat91 existed as a monomer in solution while the tyrosine phosphorylated form exists as a dimer.

10 When fractions from the glycerol gradients were analyzed by electrophoretic mobility shift analysis (Figure 17B), the peak of the phosphorylated form of Stat91 correlated well with the DNA-binding activity of Stat91. Thus only the phosphorylated dimeric Stat91 has the sequence-specific DNA recognition capacity.

15 *Stat91 Binds DNA as a Dimer.* Long or short versions of DNA binding protein can produce, respectively, a slower or a faster migrating band during gel retardation assays. Finding intermediate gel shift bands produced by mixing two different sized species provides evidence of dimerization of the DNA binding

20 proteins. Since Stat91 requires specific tyrosine phosphorylation in ligand-treated cells for its DNA binding, we sought evidence of formation of such heterodimers, first in transfected cells. An expression vector (MNC91I) encoding Stat91L, a recombinant form of Stat91 containing an additional 34 amino acid carboxyl terminal tag was generated. [The extra amino acids were encoded by a segment of

25 DNA sequence from plasmid pMNC (see Materials and Methods).] A Stat84 expression vector (MNC84) was also available (45). From somatic cell genetic experiments, mutant human cell lines (U3) are known that lack the Stat91/84 mRNA and proteins (29,30). The U3 cells were therefore separately transfected with vectors encoding Stat84 (MNC84) or Stat91L (MNC91L) or a mixture of

30 both vectors. Permanent transfectants expressing Stat84 (C84), Stat91L (C91L) or both proteins (Cmx) were isolated (Figure 18A).

Mobility shift analysis was performed with extracts from these stable cell lines (Figure 18B). Extracts of IFN- γ -treated C84 cells produced a faster migrating gel shift band than extracts of treated C91L cells. Most importantly, extracts from IFN- γ -treated Cmx cells expressing both Stat84 and Stat91L proteins formed an additional intermediate gel shift band. Anti-91, an antiserum against the C-terminal 38 amino acids of Stat91 (12) that are absent in Stat84, specifically removed the top two shift bands seen with the Cmx extracts. Anti-91, an antiserum against amino acids 609 to 716 (15) that recognizes both Stat91L and Stat84, proteins inhibited the binding of all three shift bands. Thus, the middle band formed by extracts of the Cmx cells is clearly identified as a heterodimer of Stat84 and Stat91L. We concluded that both Stat91 and Stat84 bind DNA as homodimers and, if present in the same cell, will form heterodimers.

We next wanted to detect the formation of dimers *in vitro*. When cytoplasmic or nuclear extracts of IFN- γ -treated C84 or C91L cells were mixed and analyzed (Figure 19), only the fast or slow migrating gel shift bands were observed. Thus it appeared that once formed *in vivo*, the dimers were stable. To promote the formation of protein interchange between the subunits of the dimer, a mixture of either cytoplasmic or nuclear extracts of IFN- γ -treated C84 or C91L cells were subjected mild denaturation-renaturation treatment: extracts were made 0.5 M with respect to guanidium hydrochloride for two minutes and then diluted for renaturation and subsequently used for gel retardation analysis. The formation of heterodimer was clearly detected after this treatment. When extracts from either C84 cells alone or C91L cells alone were subjected to the same treatment, the intermediate band did not form. The intermediate band was again proven by antiserum treatment to consist of Stat84/Stat91L dimer (data not shown).

This experiment defined conditions under which the dimer was stable, but also showed that dissociation and reassociation of the dimer *in vitro* was possible. Since guanidium hydrochloride is known to disrupt only non-covalent chemical

bonds, it seemed that Stat91 (or Stat84) homodimerization was mediated through non-covalent interactions.

Dimerization of Stat91 Involves Phosphotyrosyl Peptide and SH2 Interactions.

- 5 Based on the results described above, we devised a dissociation-reassociation assay in the absence of guanidium hydrochloride to explore the possible nature of interactions involved in dimer formation (Figure 20). When the short and the long forms of a homodimer are mixed with a dissociating agent (*e.g.*, a peptide containing the putative dimerization domain), the subunits of the dimer should
- 10 dissociate (in a concentration dependent fashion) due to the interaction of the agent with the dimerization domain(s) of the protein. When a specific DNA probe is subsequently added to the mixture to drive the formation of a stable protein-DNA complex, the detection of any reassociated or remaining dimers can be assayed. In the presence of low concentration of the dissociating agent, addition of DNA to
- 15 form the stable protein-DNA complex should lead to the detection of homodimers as well as heterodimers. At high concentration of the dissociating agent, subunits of the dimer may not be able to re-form and no DNA-protein complexes would be detected (Figure 20).
- 20 The Stat91 sequence contains an SH2 domain (amino acids 569 to 700, see discussion below), and we knew that Tyr-701 was the single phosphorylated tyrosine residue required for DNA binding activity (*supra*, 45). Furthermore, we have observed that phosphotyrosine at 10 mM, but not phosphoserine or phosphothreonine, could prevent the formation of Stat91-DNA complex. We
- 25 therefore sought evidence that the dimerization of Stat91 involved specific SH2-phosphotyrosine interaction using the dissociation and reassociation assay.

In order to evaluate the role of the SH2-phosphotyrosine interaction, two peptides fragments of Stat91 corresponding to segments of the SH2 and phosphotyrosing

30 domains of Stat91 were prepared: a non-phosphorylated peptide (91Y),

LDGPKGTGYIKTELI (SEQ. ID NO:18) (corresponding to amino acids 693-707), and a phosphotyrosyl peptide (91Y-p), GY*IKTE (SEQ. ID NO:19) (representing residues 700-705).

5 Activated Stat84 or Stat91L was obtained from IFN- γ -treated C84 or C91L cells and mixed in the presence of various concentrations of the peptides followed by gel mobility shift analysis. The non-phosphorylated peptide had no effect on the presence of the two gel shift bands characteristic of Stat84 or Stat91L homodimers (Figure 21, lane 2-4). In contrast, the phosphorylated peptide (91Y-p) at the
10 concentration of 4 μ M clearly promoted the exchange between the subunits of Stat84 dimers and Stat91L dimers to form heterodimers (Figure 21, lane 5). At a higher concentration (160 μ M), peptide 91Y-p but not the unphosphorylated peptide dissociated the dimers and blocked the formation of DNA protein complexes (Figure 21, lane 7).

15 When cells are treated with IFN- α both Stat91 (or 84) and Stat113 become phosphorylated (15). Antiserum to Stat113 can precipitate both Stat113 and Stat91 after IFN- α -treatment but not before, suggesting IFN- α dependent interaction of these two proteins, perhaps as a heterodimer (15).

20 In Stat113, tyr-690 in the homologous position to Tyr-701 in Stat91 is the single target residue for phosphorylation. Amino acids downstream of the affected tyrosine residue show some homology between the two proteins. We therefore prepared a phosphotyrosyl peptide of Stat113 (113Y-p), KVNQERRKY*LKHR
25 (SEQ. ID NO:20) [amino acids 681 to 694; (38)]. At concentrations similar to 91Y-p, 113Y-p also promoted the exchange of subunits between the Stat84 and Stat91L, while at a high concentration (40 μ M), 113Y-p prevented the gel shift bands almost completely (Figure 21, lane 8-10).

30 We prepared a phosphotyrosyl peptide (SrcY-p), EPQY*EEIPIYL (SEQ. ID NO:21) which is known to interact with the Src SH2 domain with a high affinity

(50). This peptide showed no effect on the Stat91 dimer formation (Figure 21, lane 11-13). Thus, it seems that Stat91 dimerization involves SH2 interaction with tyrosine residues in specific peptide sequence.

- 5 To test further the specificity of Stat91 dimerization mediated through specific-phosphotyrosyl-peptide SH2 interaction, a fusion product of glutathione-S-transferase with the Stat91-SH2 domain (GST-91SH2) was prepared (Figure 22A) and used in the *in vitro* dissociation reassociation assay. At concentrations of 0.5 to 5 μ M, the Stat91-SH2 domain promoted the formation of a heterodimer (Figure 10 22B, lanes 5-7). In contrast, neither GST alone, nor fusion products with a mutant ($R^{602} \rightarrow L^{602}$) Stat91-SH2 domain (GST 91mSH2) that renders Stat91 non-functional *in vivo*, a Stat91 SH3 domain (GST-91SH3), nor the Src SH2 domain (GST-SrcSH2), induced the exchange of subunits between the Stat84 and Stat91L homodimers (Figure 22B).

15

Discussion

- The initial sequence analysis of the Stat91 and Stat113 proteins revealed the presence of SH2 like domains (see 13,38). Further it was found that STAT 20 proteins themselves are phosphorylated on single tyrosine residues during their activation (15,31). Single amino acid mutations either removing the Stat91 phosphorylation site, Tyr-701, or converting Arg-702 to Leu in the highly conserved "pocket" region of the SH2 domain abolished the activity of Stat91 (45). Thus it seemed highly likely that one possible role of the STAT SH2 domains 25 would be to bind the phosphotyrosine residues in one of the JAK kinases.

- Since the activated STATs have phosphotyrosine residues and SH2 domains, a second suggested role for SH2 domains was in protein-protein interactions within the STAT family. By two physical criteria -- electrophoresis in native gels and 30 sedimentation on gradients -- Stat91 in untreated cells is a monomer and in treated cells is a dimer (Figures 16-18). Since phosphotyrosyl peptides from Stat91 or

Stat113 and the SH2 domain of Stat91 could efficiently promote the formation of herterodimers between Stat91L and Stat84 in a disassociation and reassociation assay, we conclude that dimerization of Stat91 involves SH2-phosphotyrosyl peptide interactions.

5

The possibility of an SH2 domain in Stat91 was indicated initially by the presence of highly conserved amino acid stretches between the Stat91 and Stat113 sequences in the 569 to 700 residue region, several of which, especially the FLLR sequence in the amino terminal end of the region, are characteristic of -SH2 domains. The C-terminal half of the SH2 domains are less well conserved in general (39); this was also true for the STAT proteins compared to other proteins, although Stat91 and Stat113 are quite similar in this region (38, 13, Figure 23). The available structures of lck, src, abl, and p85a SH2's permit identification of structurally conserved regions (SCR's), and detailed alignment of amino acid sequences of several proteins (Figure 23) is based on these.

10
15

The characteristic W (in β A1) is preceded by hydrophilic residues and is followed by hydrophobic residues in Stat91, but alignment to the W seems justified, even if the small beta sheet of which the W is part is shifted in Stat91. The three positively charged residues contributing to the phosphotyrosyl binding site are at the positions indicated as alphaA2, betaB5, and betaD5. Figure 23 shows an alignment which accomplishes this by insertions in the 'AA' and 'CD' regions. This is a different alignment from that previously suggested (38), and gives a satisfactory alignment in the (beta)D region, although, like the previous alignment, it is obviously considerably less similar to the other SH2's in the C-terminus.

20
25

This alignment suggests that the SH2 domain in the Stat91 would end in the vicinity of residue 700. In such an alignment, the Tyr-701 occurs almost immediately after the SH2 domain: a distance too short to allow an intramolecular phosphotyrosine -SH2 interaction. Since the data presented earlier strongly implicate that an SH2-phosphotyrosine interaction is involved in dimerization, such

30

an interaction is likely to be between two phospho Stat91 subunits as a reciprocal pTyr -SH2 interaction.

The apparent stability of Stat91 dimer may be due to a high association rate
5 coupled with a high dissociation rate of SH2-phosphotyrosyl peptide interactions as suggested (Felder et al., 1993, Mol. Cell Biol. 13:1449-1455) coupled with interactions between other domains of Stat91 that may contribute stability to the Stat91 dimer. Interference by homologous phosphopeptides with the -SH2-phosphotyrosine interaction would then lower stability sufficiently to allow
10 complete dissociation and heterodimerization.

The dimer formation between phospho Stat91 is the first case in eukaryotes where dimer formation is regulated by phosphorylation, and the only one thus far dependent on tyrosine phosphorylation. We anticipate that dimerization with the
15 STAT protein family will be important. It seems likely that in cells treated with IFN- α , there is Stat113-Stat91 interaction (15). This may well be mediated through SH2 and phosphotyrosyl peptide interactions as described above, leading to a complex (a probable dimer of Stat91-Stat113) which joins with a 48 kD DNA binding protein (a member of another family of DNA binding factors) to make a
20 complex capable of binding to a different DNA site. Furthermore, we have recently cloned two mouse cDNAs which encode other STAT family members that have conserved the same general structure features observed in the Stat91 and Stat113 molecules (see Example 5, *Supra*). (U.S. Application Serial No. 08/126,588, filed September 29, 1993, which is specifically incorporated herein by
25 reference in its entirety). Thus the specificity of STAT-containing complexes will almost surely be affected by which proteins are phosphorylated and then available for dimer formation.

The following is a list of references related to the above disclosure and particularly
30 to the experimental procedures and discussions. The references are numbered to correspond to like number references that appear hereinabove.

1. Lerner, A. C., Jonak, G., Cheng, Y. S., Korant, B., Knight, E. and Darnell, J. E., Jr. (1984). *Proc. Natl. Acad. Sci. USA* **81**:6733-6737; Lerner, A. C., Chaudhuri, A. and Darnell, J. E. (1986). *J. Biol. Chem.* **261**:453-459.
- 5 2. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. and Stark, G. R. (1984). *Cell* **38**:745-755.
- 10 3. Levy, D. E., Kessler, D. S., Pine, R., Reich, N. and Darnell, J. E. (1988). *Genes & Dev.* **2**:383-392.
4. Levy, D. E., Kessler, D. S., Pine, R., and Darnell, J. E. (1989). *Genes & Dev.* **3**:1362-1371.
- 15 5. Dale, T. C., Iman, A. M. A., Kerr, I. M. and Stark, G. R. (1989). *Proc. Natl. Acad. Sci.* **86**:1203-1207.
6. Kessler, D. S., Veals, S. A., Fu, X.-Y., and Levy, D. E. (1990). *Genes & Dev.* **4**:1753-1765.
- 20 7. Fu, X. -Y., Kessler, D. S., Veals, S. A., Levy, D. E. and Darnell, J. E. (1990). *Proc. Natl. Acad. Sci. USA* **87**:8555-8559.
8. Lew, D. J., Decker, T., and Darnell, J. E. (1989). *Mol. Cell. Biol.* **9**:5404-5411.
- 25 9. Decker, T., Lew, D. J., Cheng, Y.-S., Levy, D. E. and Darnell, J. E. (1989). *EMBO J.* **8**:2009-2014.
- 30 10. Decker, T., Lew, D. J., Mirkovitch, J. and Darnell, J. E., 1991. *EMBO J.* **10**:927-932.

11. Veals, S. A., Schindler, C. W., Fu, X.-Y., Leonard, D., Darnell, J. E. and Levy, D. E. (1992). *Mol. Cell. Biol.* 12.
12. Schindler, C., Fu, X.-Y., Improtta, T., Aebersold, R. and Darnell, J. E.
5 (1992). *Proc. Natl. Acad. Sci. USA* 89:7836-7389.
13. Fu, X.-Y., Schindler, C., Improtta, T., Aebersold, R. and Darnell, J. E. (1992). *Proc. Natl. Acad. Sci. USA* 89:7840-7389.
- 10 14. Lew, D. J., Decker, T. and Darnell, J. E. (1991). *Mol. Cell. Biol.* 11:182-191.
- 15 15. Schindler, C., Shuai, K., Fu, X.-Y., Prezioso, V. and Darnell, J. E. (1992). *Science* 257:809-812.
16. Garner, M. M. and Revan, A. (1981). *Nuc. Acids Res.* 9:3047-3059; Fried, A., and Crothers, D. M. (1981) *ibid* 6505-6525.
- 20 17. Celis, J. E., Justessen, J., Madsun, P. S., Lovmand, J., Ratz, G. P. and Celis, A. (1987). *Leukemia* 1:800-813.
18. Chodosh, L. A., Carthew, R. W. and Sharp, P. A. (1986). *Mol. Cell Biol.* 6:4723-4733.
- 25 19. Reich, N. and Pfeffer, L. M. (1990). *Proc. Natl. Acad. Sci. USA* 87:8761-8765.
20. Kessler, D. S. and Levy, D. (1991). *J. Biol. Chem.* 266: 23471-23476.
- 30 21. Levy, D., and Darnell, J. E. (1990). *The New Biologist* 2:923-928.

22. Aguet, J. M., Denbie, Z. and Merlin, G. (1986). *Cell* 55:273-280.
23. Kaplan, D. R., Martin-Zanca, D. and Parada, L. F. (1991). *Nature* 350:158-160; Hempstead, G., Kapland, D., Martin-Zanca, D., Parada, L. F. and Chao, M. (1991). *Nature* 350:678-683.
24. Veillette, A., Bookman, M. A., Horak, E. M., and Bolen, J. B. (1988). *Cell* 55:301-308; Rudd, C. E. et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:5190-5194.
25. Evans, R. K., Johnson, J. D. and Haley, B. E. (1986). *Proc. Natl. Acad. Sci. USA* 83:5382-5386.
26. Walaas, S. I. and Nairn, A. C. (1989). *J. of Mol. Neurosci.* 1:117-127.
28. Pellegrini, S., John, J., Shearer, M., Kerr, I.M., and Stark, G.R. (1989). *Mol. Cell. Biol.* 9:4605.
29. McKendry, R. et al. (1991). *Proc. Natl. Acad. Sci. U.S.A.* 88:11455.
30. Muller, M. et al. (1993), *EMBO. J.*
31. Shuai, K., Schindler, C., Prezioso, V., and Darnell, J.E., Jr. (1992). *Science* 258:1808.
32. Boyle, W.J., van der Geer, and Hunter, T. (1991). *Methods Enzymol.* 202:110.
33. Walaas, S.I., and Nairn, A.C. (1989). *J. Mol. Neurosci.* 1:117.
34. Khan, K.D. et al. (1993). *Proc. Natl. Acad. Sci. U.S.A.* 90:6806.

35. Qufeshi, S.A. et al. (1991). *J. Biol. Chem.* **266**:20594.
36. Kawasaki, E., (1990). In *PRC Protocols: A Guide to Methods and Applications*. Innis, M., Gelfand, D., Sinisky, J., & White, T., Eds. (San
5 Diego, CA: Academic Press), p. 119.
37. Brasier, A.R., Tata, J.E., and Habener, J.F. (1989).
Biotechniques **7**:1116.
- 10 38. Fu, X.-Y. (1992). *Cell* **70**:323-335.
39. Koch, C.A., Anderson, D. Moran, M.F., Ellis, C. and Pawson, T. (1991).
Science **252**:668-674.
- 15 40. Waksman, G. et al. (1992). *Nature* **358**:646-653.
41. Overduin, M. Rios, C.B., Mayer, B.J., Baltimore, D. and Cowburn, D.
(1992). *Cell* **70**:697-704.
- 20 42. Velazquez et al. (1992). *Cell* **70**:313.
43. Muller et al. (1993). *Nature* **366**:129-135.
44. Watling et al. (1993). *Nature* **366**:166.
- 25 45. Shuai, K., Stark, G.R., Kerr, I.M., and Darnell, J.E. (1993). *Science*
261:1744.
46. Dignam et al. (1983). *Nucl. Acids Res.* **11**:1475.
- 30 47. Davis, B.J. (1964). *Ann. N.Y. Acad. Sci.* **121**:404.

48. Bryan, J.K. (1977). *Anal. Biochem.* 78:513.
49. Amster-Choder, O. and Wright, A. (1992). *Science* 257:1395.
- 5 50. Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T.,
Haser, W.G., King, F., Roberts, T., Ratnofsky, et al. (1993). *Cell*
72:767.

10 This invention may be embodied in other forms or carried out in other ways
without departing from the spirit or essential characteristics thereof. The present
disclosure is therefore to be considered as in all respects illustrative and not
restrictive, the scope of the invention being indicated by the appended Claims, and
all changes which come within the meaning and range of equivalency are intended
to be embraced therein.

WHAT IS CLAIMED IS:

- 1 1. A receptor recognition factor implicated in the transcriptional stimulation of
2 genes in target cells in response to the binding of a specific polypeptide ligand to
3 its cellular receptor on said target cell, said receptor recognition factor having the
4 following characteristics:
 - 5 a) apparent direct interaction with the ligand-bound receptor and
6 activation of one or more transcription factors capable of binding with a specific
7 gene;
 - 8 b) an activity demonstrably unaffected by the presence or concentration
9 of second messengers;
 - 10 c) direct interaction with tyrosine kinase domains; and
 - 11 d) a perceived absence of interaction with G-proteins.
- 1 2. The receptor recognition factor of Claim 1 which is proteinaceous in
2 composition.
- 1 3. The receptor recognition factor of Claim 1 which is cytoplasmic in origin.
- 1 4. The receptor recognition factor of Claim 1 which is a polypeptide having
2 an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ
3 ID NO:10 and SEQ ID NO:12.
- 1 5. The receptor recognition factor of Claim 1 which is derived from
2 mammalian cells.
- 1 6. The receptor recognition factor of Claim 1 labeled with a detectable label.
- 1 7. The receptor recognition factor of Claim 6 wherein the label is selected
2 from enzymes, chemicals which fluoresce and radioactive elements.

- 1 8. An antibody to a receptor recognition factor, the factor to which said
2 antibody is raised having the following characteristics:
 - 3 a) apparent direct interaction with the ligand-bound receptor and
4 activation of one or more transcription factors capable of binding with a specific
5 gene;
 - 6 b) an activity demonstrably unaffected by the presence or concentration
7 of second messengers; and
 - 8 c) direct interaction with tyrosine kinase domains; and
 - 9 d) a perceived absence of interaction with G-proteins.
- 1 9. The antibody of Claim 8 which is a polyclonal antibody.
- 1 10. The antibody of Claim 8 which is a monoclonal antibody.
- 1 11. An immortal cell line that produces a monoclonal antibody according to
2 Claim 10.
- 1 12. The antibody of Claim 8 labeled with a detectable label.
- 1 13. The antibody of Claim 12 wherein the label is selected from enzymes,
2 chemicals which fluoresce and radioactive elements.
- 1 14. A DNA sequence or degenerate variant thereof, which encodes a receptor
2 recognition factor, or a fragment thereof, selected from the group consisting of:
 - 3 (A) the DNA sequence of FIGURE 1;
 - 4 (B) the DNA sequence of FIGURE 14;
 - 5 (C) the DNA sequence of FIGURE 15;
 - 6 (D) DNA sequences that hybridize to any of the foregoing DNA
7 sequences under standard hybridization conditions; and
 - 8 (E) DNA sequences that code on expression for an amino acid sequence
9 encoded by any of the foregoing DNA sequences.

1 15. A recombinant DNA molecule comprising a DNA sequence or degenerate
2 variant thereof, which encodes a receptor recognition factor, or a fragment
3 thereof, selected from the group consisting of:
4 (A) the DNA sequence of FIGURE 1;
5 (B) the DNA sequence of FIGURE 14;
6 (C) the DNA sequence of FIGURE 15;
7 (D) DNA sequences that hybridize to any of the foregoing DNA
8 sequences under standard hybridization conditions; and
9 (E) DNA sequences that code on expression for an amino acid sequence
10 encoded by any of the foregoing DNA sequences.

1 16. The recombinant DNA molecule of either of Claims 14 or 15, wherein said
2 DNA sequence is operatively linked to an expression control sequence.

1 17. The recombinant DNA molecule of Claim 16, wherein said expression
2 control sequence is selected from the group consisting of the early or late
3 promoters of SV40 or adenovirus, the lac system, the trp system, the TAC system,
4 the TRC system, the major operator and promoter regions of phage λ , the control
5 regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the
6 promoters of acid phosphatase and the promoters of the yeast α -mating factors.

1 18. A probe capable of screening for the receptor recognition factor in alternate
2 species prepared from the DNA sequence of Claim 14.

1 19. A unicellular host transformed with a recombinant DNA molecule
2 comprising a DNA sequence or degenerate variant thereof, which encodes a
3 receptor recognition factor, or a fragment thereof, selected from the group
4 consisting of:

- 5 (A) the DNA sequence of FIGURE 1;
6 (B) the DNA sequence of FIGURE 14;

- 7 (C) the DNA sequence of FIGURE 15;
8 (D) DNA sequences that hybridize to any of the foregoing DNA
9 sequences under standard hybridization conditions; and
10 (E) DNA sequences that code on expression for an amino acid sequence
11 encoded by any of the foregoing DNA sequences;
12 wherein said DNA sequence is operatively linked to an expression control
13 sequence.

1 20. The unicellular host of Claim 19 wherein the unicellular host is selected
2 from the group consisting of E. coli, Pseudomonas, Bacillus, Streptomyces,
3 yeasts, CHO, R1.1, B-W, L-M, COS 1, COS 7, BSC1, BSC40, and BMT10 cells,
4 plant cells, insect cells, and human cells in tissue culture.

1 21. A method for detecting the presence or activity of a receptor recognition
2 factor, said receptor recognition factor having the following characteristics:
3 apparent direct interaction with the ligand-bound receptor and activation of one or
4 more transcription factors capable of binding with a specific gene; an activity
5 demonstrably unaffected by the presence or concentration of second messengers;
6 direct interaction with tyrosine kinase domains; and a perceived absence of
7 interaction with G-proteins, wherein said receptor recognition factor is measured
8 by:

9 A. contacting a biological sample from a mammal in which the
10 presence or activity of said receptor recognition factor is suspected with a binding
11 partner of said receptor recognition factor under conditions that allow binding of
12 said receptor recognition factor to said binding partner to occur; and

13 B. detecting whether binding has occurred between said receptor
14 recognition factor from said sample and the binding partner;
15 wherein the detection of binding indicates that presence or activity of said
16 receptor recognition factor in said sample.

1 22. A method for detecting the presence and activity of a polypeptide ligand
2 associated with a given invasive stimulus in mammals comprising detecting the
3 presence or activity of a receptor recognition factor according to the method of
4 Claim 21, wherein detection of the presence or activity of the receptor recognition
5 factor indicates the presence and activity of a polypeptide ligand associated with a
6 given invasive stimulus in mammals.

1 23. The method of Claim 22 wherein said invasive stimulus is an infection.

1 24. The method of Claim 22 wherein said invasive stimulus is selected from
2 the group consisting of viral infection, protozoan infection, tumorous mammalian
3 cells, and toxins.

1 25. A method for detecting the binding sites for a receptor recognition factor,
2 said receptor recognition factor having the following characteristics:

3 apparent direct interaction with the ligand-bound receptor and activation of
4 one or more transcription factors capable of binding with a specific gene;

5 an activity demonstrably unaffected by the presence or concentration of
6 second messengers;

7 direct interaction with tyrosine kinase domains; and

8 a perceived absence of interaction with G-proteins; wherein the binding
9 sites for said receptor recognition factor are measured by:

10 A. placing a labeled receptor recognition factor sample in
11 contact with a biological sample from a mammal in which binding sites for said
12 receptor recognition factor are suspected;

13 B. examining said biological sample in binding studies for the
14 presence of said labeled receptor recognition factor;

15 wherein the presence of said labeled recognition factor indicates a binding
16 site for a receptor recognition factor.

- 1 26. A method of testing the ability of a drug or other entity to modulate the
2 activity of a receptor recognition factor which comprises
- 3 A. culturing a colony of test cells which has a receptor for the
4 receptor recognition factor in a growth medium containing the receptor recognition
5 factor;
- 6 B. adding the drug under test; and
- 7 C. measuring the reactivity of said receptor recognition factor with the
8 receptor on said colony of test cells,
- 9 wherein said receptor recognition factor has the following characteristics:
- 10 a) apparent direct interaction with the ligand-bound receptor and
11 activation of one or more transcription factors capable of binding with a specific
12 gene;
- 13 b) an activity demonstrably unaffected by the presence or concentration
14 of second messengers;
- 15 c) direct interaction with tyrosine kinase domains; and
- 16 d) a perceived absence of interaction with G-proteins.

- 1 27. An assay system for screening drugs and other agents for ability to
2 modulate the production of a receptor recognition factor, comprising:
- 3 A. culturing an observable cellular test colony inoculated with a drug
4 or agent;
- 5 B. harvesting a supernatant from said cellular test colony; and
- 6 C. examining said supernatant for the presence of said receptor
7 recognition factor wherein an increase or a decrease in a level of said receptor
8 recognition factor indicates the ability of a drug to modulate the activity of said
9 receptor recognition factor, said receptor recognition factor having the following
10 characteristics:
- 11 a) apparent direct interaction with the ligand-bound receptor and
12 activation of one or more transcription factors capable of binding with a specific
13 gene;

- 14 b) an activity demonstrably unaffected by the presence or concentration
15 of second messengers;
16 c) direct interaction with tyrosine kinase domains; and
17 d) a perceived absence of interaction with G-proteins.

- 1 28. A test kit for the demonstration of a receptor recognition factor in a
2 eukaryotic cellular sample, comprising:
3 A. a predetermined amount of a detectably labelled specific binding
4 partner of a receptor recognition factor, said receptor recognition factor having the
5 following characteristics: apparent direct interaction with the ligand-bound receptor
6 and activation of one or more transcription factors capable of binding with a
7 specific gene; an activity demonstrably unaffected by the presence or concentration
8 of second messengers; direct interaction with tyrosine kinase domains; and a
9 perceived absence of interaction with G-proteins;
10 B. other reagents; and
11 C. directions for use of said kit.

29. A test kit for demonstrating the presence of a receptor recognition factor in
a eukaryotic cellular sample, comprising:

- A. a predetermined amount of a receptor recognition factor, said
receptor recognition factor having the following characteristics: apparent direct
interaction with the ligand-bound receptor and activation of one or more
transcription factors capable of binding with a specific gene; an activity
demonstrably unaffected by the presence or concentration of second messengers;
direct interaction with tyrosine kinase domains; and a perceived absence of
interaction with G-proteins;
 B. a predetermined amount of a specific binding partner of said
receptor recognition factor;
 C. other reagents; and
 D. directions for use of said kit;

wherein either said receptor recognition factor or said specific binding partner are detectably labelled.

1 30. The test kit of Claim 28 or 29 wherein said labeled immunochemically
2 reactive component is selected from the group consisting of polyclonal antibodies
3 to the receptor recognition factor, monoclonal antibodies to the receptor
4 recognition factor, fragments thereof, and mixtures thereof.

1 31. A method of preventing and/or treating cellular debilitations, derangements
2 and/or dysfunctions and/or other disease states in mammals, comprising
3 administering to a mammal a therapeutically effective amount of a material
4 selected from the group consisting of a receptor recognition factor, an agent
5 capable of promoting the production and/or activity of said receptor recognition
6 factor, an agent capable of mimicking the activity of said receptor recognition
7 factor, an agent capable of inhibiting the production of said receptor recognition
8 factor, and mixtures thereof, or a specific binding partner thereto, said receptor
9 recognition factor having the following characteristics:
10 a) apparent direct interaction with the ligand-bound receptor and
11 activation of one or more transcription factors capable of binding with a specific
12 gene;
13 b) an activity demonstrably unaffected by the presence or concentration
14 of second messengers;
15 c) direct interaction with tyrosine kinase domains; and
16 d) a perceived absence of interaction with G-proteins.

1 32. The method of Claim 31 wherein said disease states include chronic viral
2 hepatitis, hairy cell leukemia, and tumorous conditions.

1 33. The method of Claim 31 wherein said receptor recognition factor is
2 administered to modulate the course of therapy where interferon is being
3 administered as the primary therapeutic agent.

1 34. The method of Claim 31 wherein said receptor recognition factor is
2 administered to modulate the course of therapy where interferon is being co-
3 administered with one or more additional therapeutic agents.

1 35. A pharmaceutical composition for the treatment of cellular debilitation,
2 derangement and/or dysfunction in mammals, comprising:
3 A. a therapeutically effective amount of a material selected from
4 the group consisting of a receptor recognition factor, an agent capable of
5 promoting the production and/or activity of said receptor recognition factor, an
6 agent capable of mimicking the activity of said receptor recognition factor, an
7 agent capable of inhibiting the production of said receptor recognition factor, and
8 mixtures thereof, or a specific binding partner thereto, said receptor recognition
9 factor having the following characteristics: apparent direct interaction with the
10 ligand-bound receptor and activation of one or more transcription factors capable
11 of binding with a specific gene; an activity demonstrably unaffected by the
12 presence or concentration of second messengers; direct interaction with tyrosine
13 kinase domains; and a perceived absence of interaction with G-proteins; and
14 B. a pharmaceutically acceptable carrier.

1 36. A receptor recognition factor implicated in the transcriptional stimulation of
2 genes in target cells in response to the binding of a specific polypeptide ligand to
3 its cellular receptor on said target cell, said receptor recognition factor having the
4 following properties:
5 a) it is present in cytoplasm;
6 b) it undergoes tyrosine phosphorylation upon treatment of cells with
7 IFN α ;
8 c) it activates transcription of an interferon stimulated gene;
9 d) it stimulates either an ISRE-dependent or a gamma activated site
10 (GAS)-dependent transcription in vivo;
11 e) it interacts with IFN α cellular receptors, and

12 f) it undergoes nuclear translocation upon stimulation of the IFN cellular
13 receptors with IFN α .

1 37. A receptor recognition factor implicated in the transcriptional stimulation of
2 genes in target cells in response to the binding of an interferon or interferon-
3 related polypeptide ligand to its cellular receptor on said target cell, said receptor
4 recognition factor having the following properties:

5 a) it is present in vivo in mammalian cytoplasm before activation of
6 cellular IFN receptors;

7 b) it contains tyrosine sites that are phosphorylated in response to IFN
8 stimulation of IFN receptors;

9 c) it has a molecular weight selected from the group consisting of 48kD,
10 84kD, 91kD and 113kD, or an amino acid sequence selected from the group
11 consisting of SEQ ID NO:10 and SEQ ID NO:12, and

12 d) when phosphorylated, it recognizes an ISRE in the cell nucleus.

1 38. The receptor recognition factor of either of Claims 36 or 37 in
2 phosphorylated form.

1 39. An antibody which recognizes a phosphorylated ISGF3 polypeptide or a
2 fragment thereof in phosphorylated form.

1 40. An antibody produced by injecting a substantially immunocompetent host
2 with an antibody-producing effective amount of an ISGF3 polypeptide, and
3 harvesting said antibody, said ISGF3 polypeptide having the following properties:

4 a) it has a molecular weight of about 48kD, 84Kd, 91 Kd or 113kD or an
5 amino acid sequence selected from the group consisting of SEQ ID NO:10 and
6 SEQ ID NO:12;

7 b) it can be isolated from mammalian cytoplasm;

8 c) it contains tyrosine residues that are subject to phosphorylation in vivo
9 upon treatment of cells with IFN α ;

- 10 d) it can activate transcription of an interferon stimulated gene in vivo;
11 e) it can stimulate ISRE-dependent transcription in vivo;
12 f) it can interact with IFN α cellular receptors, and
13 g) it can undergo nuclear translocation upon stimulation of IFN cellular
14 receptors with IFN α .

1 41. The antibody of either of Claims 39 or 40 which is monoclonal.

1 42. The antibody of either of Claims 39 or 40 which is polyclonal.

1 43. A recombinant virus transformed with the DNA molecule, or a derivative
2 or fragment thereof, in accordance with Claim 14.

1 44. A recombinant virus transformed with the DNA molecule, or a derivative
2 or fragment thereof, in accordance with Claim 15.

1 45. A method of enhancing IFN α activity in a mammal in need of such
2 treatment, comprising administering to said mammal an effective amount of a
3 compound which (a) enhances the phosphorylation of intracellular ISGF3 proteins
4 to form ISGF3-protein phosphates, or (b) inhibits the activity of a phosphatase
5 enzyme which would otherwise reduce the level of phosphorylated ISGF3 proteins.

1 46. A method of treating (a) chronic viral hepatitis or (b) hairy cell leukemia,
2 in a mammal in need of such treatment, comprising administering to said mammal
3 an effective amount of a compound which (a) enhances the phosphorylation of
4 ISGF3 proteins, or (b) decreases the level of phosphate removal from
5 phosphorylated ISGF3 proteins.

1 47. The method of Claim 45 wherein the activity of exogenous IFN α is
2 enhanced.

- 1 48. The method of Claim 45 wherein the activity of endogenous IFN α is
2 enhanced.
- 1 49. The method of Claim 47 wherein the compound and IFN α are administered
2 concurrently to the mammal in need of such treatment.
- 1 50. A method of determining the interferon-related pharmacological activity of
2 a compound comprising:
3 administering the compound to a mammal;
4 determining the level of phosphorylated ISGF3 proteins present; and
5 comparing the level of ISGF3 protein-phosphate to a standard.
- 1 51. In a method of treating hepatitis or leukemia in a mammal, wherein IFN α
2 is administered in an amount effective for treating such hepatitis or leukemia, the
3 improvement comprising administering to said mammal an ISGF3 protein or a
4 derivative thereof in an amount effective for enhancing the activity of said IFN α .
- 1 52. The method of Claim 51 wherein a derivative of said ISGF3 protein is
2 administered.
- 1 53. The method of Claim 51 wherein an ISGF3 protein is administered, having
2 a molecular weight of about 48 kD, 84kD, 91kD or 113kD.
- 1 54. The method of Claim 52 wherein the derivative is a phosphorylated ISGF3
2 protein.
- 1 55. The recombinant DNA molecule of Claim 16 comprising plasmid pGEX-
2 3X, clone E3 or plasmid pGEX-3X, clone E4.
- 1 56. An antisense nucleic acid against a receptor recognition factor mRNA
2 comprising a nucleic acid sequence hybridizing to said mRNA.

- 1 57. The antisense nucleic acid of Claim 56 which is RNA.
- 1 58. The antisense nucleic acid of Claim 56 which is DNA.
- 1 59. The antisense nucleic acid of Claim 56 which binds to the initiation codon
2 of any of said mRNAs.
- 1 60. A recombinant DNA molecule having a DNA sequence which, on
2 transcription, produces an antisense ribonucleic acid against a receptor recognition
3 factor mRNA, said antisense ribonucleic acid comprising an nucleic acid sequence
4 capable of hybridizing to said mRNA.
- 1 61. A receptor recognition factor-producing cell line transfected with the
2 recombinant DNA molecule of Claim 60.
- 1 62. A method for creating a cell line which exhibits reduced expression of a
2 receptor recognition factor, comprising transfecting a recognition factor-producing
3 cell line with a recombinant DNA molecule of claim 60.
- 1 63. A ribozyme that cleaves receptor recognition factor mRNA.
- 1 64. The ribozyme of Claim 63 which is a Tetrahymena-type ribozyme.
- 1 65. The ribozyme of Claim 63 which is a Hammerhead-type ribozyme.
- 1 66. A recombinant DNA molecule having a DNA sequence which, upon
2 transcription, produces the ribozyme of claim 63.
- 1 67. A receptor recognition factor-producing cell line transfected with the
2 recombinant DNA molecule of claim 66.

- 1 68. A method for creating a cell line which exhibits reduced expression of a
- 2 receptor recognition factor, comprising transfecting a recognition factor-producing
- 3 cell line with the recombinant DNA molecule of claim 63.

ABSTRACT

Receptor recognition factors exist that recognizes the specific cell receptor to which a specific ligand has been bound, and that may thereby signal and/or initiate the binding of the transcription factor to the DNA site. The receptor recognition factor is in one instance, a part of a transcription factor, and also may interact with other transcription factors to cause them to activate and travel to the nucleus for DNA binding. The receptor recognition factor appears to be second-messenger-independent in its activity, as overt perturbations in second messenger concentrations are of no effect. The concept of the invention is illustrated by the results of studies conducted with interferon (IFN)-stimulated gene transcription, and particularly, the activation caused by both IFN α and IFN γ . Specific DNA and amino acid sequences for various human and murine receptor recognition factors are provided, as are polypeptide fragments of two of the ISGF-3 genes, and antibodies have also been prepared and tested. The polypeptides confirm direct involvement of tyrosine kinase in intracellular message transmission. Numerous diagnostic and therapeutic materials and utilities are also disclosed.

1
 met ala gln trp glu met leu gln
 ACTGCAACCCTAATCAGAGCCCAA ATG GCG CAG TGG GAA ATG CTG CAG
 10 20
 asn leu asp ser pro phe gln asp gln leu his gln leu tyr ser
 AAT CTT GAC AGC CCC TTT CAG GAT CAG CTG CAC CAG CTT TAC TCG
 30
 his ser leu leu pro val asp ile arg gln tyr leu ala val trp
 CAC AGC CTC CTG CCT GTG GAC ATT CGA CAG TAC TTG GCT GTC TGG
 40 50
 ile glu asp gln asn trp gln glu ala ala leu gly ser asp asp
 ATT GAA GAC CAG AAC TGG CAG GAA GCT GCA CTT GGG AGT GAT GAT
 60
 ser lys ala thr met leu phe phe his phe leu asp gln leu asn
 TCC AAG GCT ACC ATG CTA TTC TTC CAC TTC TTG GAT CAG CTG AAC
 70 80
 tyr glu cys gly arg cys ser gln asp pro glu ser leu leu leu
 TAT GAG TGT GGC CGT TGC AGC CAG GAC CCA GAG TCC TTG TTG CTG
 90
 gln his asn leu arg lys phe cys arg asp ile gln pro phe ser
 CAG CAC AAT TTG CGG AAA TTC TGC CGG GAC ATT CAG CCC TTT TCC
 100 110
 gln asp pro thr gln leu ala glu met ile phe asn leu leu leu
 CAG GAT CCT ACC CAG TTG GCT GAG ATG ATC TTT AAC CTC CTT CTG
 120
 glu glu lys arg ile leu ile gln ala gln arg ala gln leu glu
 GAA GAA AAA AGA ATT TTG ATC CAG GCT CAG AGG GCC CAA TTG GAA
 130 140
 gln gly glu pro val leu glu thr pro val glu ser gln gln his
 CAA GGA GAG CCA GTT CTC GAA ACA CCT GTG GAG AGC CAG CAA CAT
 150
 glu ile glu ser arg ile leu asp leu arg ala met met glu lys
 GAG ATT GAA TCC CGG ATC CTG GAT TTA AGG GCT ATG ATG GAG AAG
 160 170
 leu val lys ser ile ser gln leu lys asp gln gln asp val phe
 CTG GTA AAA TCC ATC AGC CAA CTG AAA GAC CAG CAG GAT GTC TTC
 180

Figure 1A

cys phe arg tyr lys ile gln ala lys gly lys thr pro ser leu
 TGC TTC CGA TAT AAG ATC CAG GCC AAA GGG AAG ACA CCC TCT CTC
 190 200
 asp pro his gln thr lys glu gln lys ile leu gln glu thr leu
 GAC CCC CAT CAG ACC AAA GAG CAG AAG ATT CTG CAG GAA ACT CTC
 210
 asn glu leu asp lys arg arg lys glu val leu asp ala ser lys
 AAT GAA CTG GAC AAA AGG AGA AAG GAG GTG CTG GAT GCC TCC AAA
 220 230
 ala leu leu gly arg leu thr thr leu ile glu leu leu leu pro
 GCA CTG CTA GGC CGA TTA ACT ACC CTA ATC GAG CTA CTG CTG CCA
 240
 lys leu glu glu trp lys ala gln gln gln lys ala cys ile arg
 AAG TTG GAG GAG TGG AAG GCC CAG CAG CAA AAA GCC TGC ATC AGA
 250 260
 ala pro ile asp his gly leu glu gln leu glu thr trp phe thr
 GCT CCC ATT GAC CAC GGG TTG GAA CAG CTG GAG ACA TGG TTC ACA
 270
 ala gly ala lys leu leu phe his leu arg gln leu leu lys glu
 GCT GGA GCA AAG CTG TTG TTT CAC CTG AGG CAG CTG CTG AAG GAG
 280 290
 leu lys gly leu ser cys leu val ser tyr gln asp asp pro leu
 CTG AAG GGA CTG AGT TGC CTG GTT AGC TAT CAG GAT GAC CCT CTG
 300
 thr lys gly val asp leu arg asn ala gln val thr glu leu leu
 ACC AAA GGG GTG GAC CTA CGC AAC GCC CAG GTC ACA GAG TTG CTA
 310 320
 gln arg leu leu his arg ala phe val val glu thr gln pro cys
 CAG CGT CTG CTC CAC AGA GCC TTT GTG GTA GAA ACC CAG CCC TGC
 330
 met pro gln thr pro his arg pro leu ile leu lys thr gly ser
 ATG CCC CAA ACT CCC CAT CGA CCC CTC ATC CTC AAG ACT GGC AGC
 340 350
 lys phe thr val arg thr arg leu leu val arg leu gln glu gly
 AAG TTC ACC GTC CGA ACA AGG CTG CTG GTG AGA CTC CAG GAA GGC
 360
 asn glu ser leu thr val glu val ser ile asp arg asn pro pro
 AAT GAG TCA CTG ACT GTG GAA GTC TCC ATT GAC AGG AAT CCT CCT
 370 380
 gln leu gln gly phe arg lys phe asn ile leu thr ser asn gln
 CAA TTA CAA GGC TTC CGG AAG TTC AAC ATT CTG ACT TCA AAC CAG
 390
 lys thr leu thr pro glu lys gly gln ser gln gly leu ile trp

Figure 1B

```

AAA ACT TTG ACC CCC GAG AAG GGG CAG AGT CAG GGT TTG ATT TGG

      400                                410
asp phe gly tyr leu thr leu val glu gln arg ser gly gly ser
GAC TTT GGT TAC CTG ACT CTG GTG GAG CAA CGT TCA GGT GGT TCA

      420
gly lys gly ser asn lys gly pro leu gly val thr glu glu leu
GGA AAG GGC AGC AAT AAG GGG CCA CTA GGT GTG ACA GAG GAA CTG

      430                                440
his ile ile ser phe thr val lys tyr thr tyr gln gly leu lys
CAC ATC ATC AGC TTC ACG GTC AAA TAT ACC TAC CAG GGT CTG AAG

      450
gln glu leu lys thr asp thr leu pro val val ile ile ser asn
CAG GAG CTG AAA ACG GAC ACC CTC CCT GTG GTG ATT ATT TCC AAC

      460                                470
met asn gln leu ser ile ala trp ala ser val leu trp phe asn
ATG AAC CAG CTC TCA ATT GCC TGG GCT TCA GTT CTC TGG TTC AAT

      480
leu leu ser pro asn leu gln asn gln gln phe phe ser asn pro
TTG CTC AGC CCA AAC CTT CAG AAC CAG CAG TTC TTC TCC AAC CCC

      490                                500
pro lys ala pro trp ser leu leu gly pro ala leu ser trp gln
CCC AAG GCC CCC TGG AGC TTG CTG GGC CCT GCT CTC AGT TGG CAG

      510
phe ser ser tyr val gly arg gly leu asn ser asp gln leu ser
TTC TCC TCC TAT GTT GGC CGA GGC CTC AAC TCA GAC CAG CTG AGC

      520                                530
met leu arg asn lys leu phe gly gln asn cys arg thr glu asp
ATG CTG AGA AAC AAG CTG TTC GGG CAG AAC TGT AGG ACT GAG GAT

      540
pro leu leu ser trp ala asp phe thr lys arg glu ser pro pro
CCA TTA TTG TCC TGG GCT GAC TTC ACT AAG CGA GAG AGC CCT CCT

      550                                560
gly lys leu pro phe trp thr trp leu asp lys ile leu glu leu
GGC AAG TTA CCA TTC TGG ACA TGG CTG GAC AAA ATT CTG GAG TTG

      570
val his asp his leu lys asp leu trp asn asp gly arg ile met
GTA CAT GAC CAC CTG AAG GAT CTC TGG AAT GAT GGA CGC ATC ATG

      580                                590
gly phe val ser arg ser gln glu arg arg leu leu lys lys thr
GGC TTT GTG AGT CGG AGC CAG GAG CGC CGG CTG CTG AAG AAG ACC

      600
met ser gly thr phe leu leu arg phe ser glu ser ser glu gly
ATG TCT GGC ACC TTT CTA CTG CGC TTC AGT GAA TCG TCA GAA GGG

```

Figure 1C

610
gly ile thr cys ser trp val glu his gln asp asp asp lys val
GGC ATT ACC TGC TCC TGG GTG GAG CAC CAG GAT GAT GAC AAG GTG

620
leu ile tyr ser val gln pro tyr thr lys glu val leu gln ser
CTC ATC TAC TCT GTG CAA CCG TAC ACG AAG GAG GTG CTG CAG TCA

630
leu pro leu thr glu ile ile arg his tyr gln leu leu thr glu
CTC CCG CTG ACT GAA ATC ATC CGC CAT TAC CAG TTG CTC ACT GAG

640
glu asn ile pro glu asn pro leu arg phe leu tyr pro arg ile
GAG AAT ATA CCT GAA AAC CCA CTG CGC TTC CTC TAT CCC CGA ATC

650
pro arg asp glu ala phe gly cys tyr tyr gln glu lys val asn
CCC CGG GAT GAA GCT TTT GGG TGC TAC TAC CAG GAG AAA GTT AAT

660
leu gln glu arg arg lys tyr leu lys his arg leu ile val val
CTC CAG GAA CGG AGG AAA TAC CTG AAA CAC AGG CTC ATT GTG GTC

670
ser asn arg gln val asp glu leu gln gln pro leu glu leu lys
TCT AAT AGA CAG GTG GAT GAA CTG CAA CAA CCG CTG GAG CTT AAG

680
pro glu pro glu leu glu ser leu glu leu glu leu gly leu val
CCA GAG CCA GAG CTG GAG TCA TTA GAG CTG GAA CTA GGG CTG GTG

690
pro glu pro glu leu ser leu asp leu glu pro leu leu lys ala
CCA GAG CCA GAG CTC AGC CTG GAC TTA GAG CCA CTG CTG AAG GCA

700
gly leu asp leu gly pro glu leu glu ser val leu glu ser thr
GGG CTG GAT CTG GGG CCA GAG CTA GAG TCT GTG CTG GAG TCC ACT

710
leu glu pro val ile glu pro thr leu cys met val ser gln thr
CTG GAG CCT GTG ATA GAG CCC ACA CTA TGC ATG GTA TCA CAA ACA

720
val pro glu pro asp gln gly pro val ser gln pro val pro glu
GTG CCA GAG CCA GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG

730
pro asp leu pro cys asp leu arg his leu asn thr glu pro met
CCA GAT TTG CCC TGT GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG

740
glu ile phe arg asn cys val lys ile glu glu ile met pro asn
GAA ATC TTC AGA AAC TGT GTA AAG ATT GAA GAA ATC ATG CCG AAT

750
val pro glu pro asp gln gly pro val ser gln pro val pro glu
GTG CCA GAG CCA GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG

760
pro asp leu pro cys asp leu arg his leu asn thr glu pro met
CCA GAT TTG CCC TGT GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG

770
glu ile phe arg asn cys val lys ile glu glu ile met pro asn
GAA ATC TTC AGA AAC TGT GTA AAG ATT GAA GAA ATC ATG CCG AAT

780
val pro glu pro asp gln gly pro val ser gln pro val pro glu
GTG CCA GAG CCA GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG

790
pro asp leu pro cys asp leu arg his leu asn thr glu pro met
CCA GAT TTG CCC TGT GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG

800
glu ile phe arg asn cys val lys ile glu glu ile met pro asn
GAA ATC TTC AGA AAC TGT GTA AAG ATT GAA GAA ATC ATG CCG AAT

810
val pro glu pro asp gln gly pro val ser gln pro val pro glu
GTG CCA GAG CCA GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG

Figure 1D

020 030
gly asp pro leu leu ala gly gln asn thr val asp glu val tyr
GGT GAC CCA CTG TTG GCT GGC CAG AAC ACC GTG GAT GAG GTT TAC

040
val ser arg pro ser his phe tyr thr asp gly pro leu met pro
GTC TCC CGC CCC AGC CAC TTC TAC ACT GAT GGA CCC TTG ATG CCT

050 051
ser asp phe AM
TCT GAC TTC TAG GAACACATTTCTCTGTTCTTTTCATATCTCTTTGCCCTTCCTA
CTCCTCATAGCATGATATTGTTCTCCAAGGATGGGAATCAGGCATGTGTCCCTTCCAAGC
TGTGTTAACTGTTCAAACCTCAGGCCTGTGTGACTCCATTGGGGTGAGAGGTGAAGCATA
ACATGGGTACAGAGGGGACAAACATGAATCAGAACAGATGCTGAGCCATAGGTCTAAATA
GGATCCTGGAGGGCTGCCTGCTGTGCTGGGAGGTATAGGGGTCTGGGGGCAGGCCAGGGC
AGTTGACAGGTACTTGGAGGGCTCAGGGCAGTGGCTTCTTCCAGTATGGAAGGATTTCA
ACATTTTAATAGTTGGTTAGGCTAAACTGGTGCATACTGGCATTGGCCTTGGTGGGGAGC
ACAGACACAGGATAGGACTCCATTTCTTTCTTCCATTCCTTCATGTCTAGGATAACTTGC
TTTCTTCTTTCTTTACTCCTGGCTCAAGCCCTGAATTTCTTCTTTTCTTCTGCAAGGGTTG
AGAGCTTTCTGCCTTAGCCTACCATGTGAAACTCTACCCCTGAAGAAAGGGATGGATAGGA
AGTAGACCTCTTTTCTTACCAGTCTCCTCCCTACTCTGCCCCCTAAGCTGGCTGTACC
TGTTCTCCCCCATAAAAATGATCCTGCCAATCTAAAAAAAAAA

Figure 1E

ATTAAACCTCTCGCCGAGCCCTCCGCAGACTCTGCGCCGAAAGTTTCATTTGCTGTATGCCATCCTCGA

GAGCTGTCTAGGTTAACGTTTCGCACCTCTGTGTATATAACCTCGACAGTCTTGGCACCTAACGTGCTGTGCG

Met Ser Gln Trp
TAGCTGCTCCTTTGGTTGAATCCCCAGGCCCTTGTGGGGCACAAGGTGGCAGG ATG TCT CAG TGG

Tyr Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu Glu Gln Val His Gln Leu Tyr
TAC GAA CTT CAG CAG CTT GAC TCA AAA TTC CTG GAG CAG GTT CAC CAG CTT TAT

Asp Asp Ser Phe Pro Met Glu Ile Arg Gln Tyr Leu Ala Gln Trp Leu Glu Lys
GAT GAC AGT TTT CCC ATG GAA ATC AGA CAG TAC CTG GCA CAG TGG TTA GAA AAG

Gln Asp Trp Glu His Ala Ala Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His
CAA GAC TGG GAG CAC GCT GCC AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT

Asp Leu Leu Ser Gln Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn
GAC CTC CTG TCA CAG CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT AAC

Phe Leu Leu Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe
TTC TTG CTA CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG GAT AAT TTT

Gln Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys Glu Glu
CAG GAA GAC CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC TGT CTG AAG GAA GAA

Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Ser Gly Asn Ile
AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT CAG GCT CAG TCG GGG AAT ATT

Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp Ser Lys Val Arg Asn
CAG AGC ACA GTG ATG TTA GAC AAA CAG AAA GAG CTT GAC AGT AAA GTC AGA AAT

Val Lys Asp Lys Val Met Cys Ile Glu His Glu Ile Lys Ser Leu Glu Asp Leu
GTG AAG GAC AAG GTT ATG TGT ATA GAG CAT GAA ATC AAG AGC CTG GAA GAT TTA

Gln Asp Glu Tyr Asp Phe Lys Cys Lys Thr Leu Gln Asn Arg Glu His Glu Thr
CAA GAT GAA TAT GAC TTC AAA TGC AAA ACC TTG CAG AAC AGA GAA CAC GAG ACC

Asn Gly Val Ala Lys Ser Asp Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met
AAT GGT GTG GCA AAG AGT GAT CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG

Tyr Leu Met Leu Asp Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu
TAT TTA ATG CTT GAC AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG TTG

Leu Asn Val Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu
CTG AAT GTC ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA CTA GTG GAG

Trp Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
TGG AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG CCC AAT GCT TGC TTG

Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln Gln Val Arg
GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG GAG AGT CTG CAG CAA GTT CGG

Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr Thr Tyr Glu His Asp
CAG CAG CTT AAA AAG TTG GAG GAA TTG GAA CAG AAA TAC ACC TAC GAA CAT GAC

Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp Arg Thr Phe Ser Leu Phe Gln
CCT ATC ACA AAA AAC AAA CAA GTG TTA TGG GAC CGC ACC TTC AGT CTT TTC CAG

Figure 2A

Gln Leu Ile Gln Ser Ser Phe Val Val Glu Arg Gln Pro Cys Met Pro Thr His
 CAG CTC ATT CAG AGC TCG TTT GTG GTG GAA AGA CAG CCC TGC ATG CCA ACG CAC

Pro Gln Arg Pro Leu Val Leu Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg
 CCT CAG AGG CCG CTG GTC TTG AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA

Leu Leu Val Lys Leu Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe
 CTG TTG GTG AAA TTG CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA TTT

Asp Lys Asp Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile
 GAT AAA GAT GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG TTC AAC ATT

Leu Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser Leu
 TTG GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC ACC AAT GGC AGT CTG

Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala Gly Thr Arg
 GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA CAG AAA AAT GCT GGC ACC AGA

Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His Ser Leu Ser Phe Glu
 ACG AAT GAG GGT CCT CTC ATC GTT ACT GAA GAG CTT CAC TCC CTT AGT TTT GAA

Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp Leu Glu Thr Thr Ser Leu Pro
 ACC CAA TTG TGC CAG CCT GGT TTG GTA ATT GAC CTC GAG ACG ACC TCT CTG CCC

Val Val Val Ile Ser Asn Val Ser Gln Leu Pro Ser Gly Trp Ala Ser Ile Leu
 GTT GTG GTG ATC TCC AAC GTC AGC CAG CTC CCG AGC GGT TGG GCC TCC ATC CTT

Trp Tyr Asn Met Leu Val Ala Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro
 TGG TAC AAC ATG CTG GTG GCG GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA

Pro Cys Ala Arg Trp Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser
 CCA TGT GCA CGA TGG GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT TCT

Val Thr Lys Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly Glu Lys Leu
 GTC ACC AAA AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA GAG AAG CTT

Leu Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe Cys Lys
 CTT GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG ACG AGG TTT TGT AAG

Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp Leu Trp Ile Glu Ser Ile Leu
 GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG CTT TGG ATT GAA AGC ATC CTA

Glu Leu Ile Lys Lys His Leu Leu Pro Leu Trp Asn Asp Gly Cys Ile Met Gly
 GAA CTC ATT AAA AAA CAC CTG CTC CCT CTC TGG AAT GAT GGG TGC ATC ATG GGC

Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu Leu Lys Asp Gln Gln Pro Gly Thr
 TTC ATC AGC AAG GAG CGA GAG CGT GCC CTG TTG AAG GAC CAG CAG CCG GGG ACC

Phe Leu Leu Arg Phe Ser Glu Ser Ser Arg Glu Gly Ala Ile Thr Phe Thr Trp
 TTC CTG CTG CGG TTC AGT GAG AGC TCC CGG GAA GGG GCC ATC ACA TTC ACA TGG

Val Glu Arg Ser Gln Asn Gly Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr
 GTG GAG CGG TCC CAG AAC GGA GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC

Thr Lys Lys Glu Leu Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys
 ACG AAG AAA GAA CTT TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC AAA

Val Met Ala Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn
 GTC ATG GCT GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG TAT CCA AAT

Figure 2B

Ile Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu Ala Pro
 ATT GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG CCA AAG GAA GCA CCA

 Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr Gly Tyr Ile Lys Thr Glu Leu
 GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT GGA TAT ATC AAG ACT GAG TTG

 Ile Ser Val Ser Glu Val His Pro Ser Arg Leu Gln Thr Thr Asp Asn Leu Leu
 ATT TCT GTG TCT GAA GTT CAC CCT TCT AGA CTT CAG ACC ACA GAC AAC CTG CTC

 Pro Met Ser Pro Glu Glu Phe Asp Glu Val Ser Arg Ile Val Gly Ser Val Glu
 CCC ATG TCT CCT GAG GAG TTT GAC GAG GTG TCT CGG ATA GTG GGC TCT GTA GAA

 Phe Asp Ser Met Met Asn Thr Val
 TTC GAC AGT ATG ATG AAC ACA GTA TAGAGCATGAATTTTTTTCATCTTCTCTGGCGACAGTTT

 TCCTTCTCATCTGTGATTCCCTCCTGCTACTCTGTTCCCTTCACATCCTGTGTTTCTAGGGAATGAAAGAA

 AGGCCAGCAAATTCGCTGCAACCTGTTGATAGCAAGTGAATTTTTCTCTAACTCAGAAACATCAGTTACTC

 TGAAGGGCATCATGCATCTTACTGAAGGTAAAAATTGAAAGGCATTCTCTGAAGAGTGGGTTTCACAAGTGA

 AAAACATCCAGATACACCCAAAGTATCAGGACGAGAATGAGGGTCCTTTGGGAAAGGAGAAGTTAAGCAAC

 ATCTAGCAAATGTTATGCATAAAGTCAGTGCCCAACTGTTATAGGTTGTTGGATAAATCAGTGGTTATTTA

 GGGAACTGCTTGACGTAGGAACGGTAAATTTCTGTGGGAGAATTCTTACATGTTTTCTTTGCTTTAAGTGT

 AACTGGCAGTTTTCCATTGGTTTACCTGTGAAATAGTTCAAAGCCAAGTTTATATACAATTATATCAGTCC

 TCTTTCAAAGGTAGCCATCATGGATCTGGTAGGGGAAAATGTGTATTTTATTACATCTTTCACATTGGCT

 ATTTAAAGACAAAGACAAATTTCTGTTTCTTGAGAAGAGAATATTAGCTTTACTGTTTGTATGGCTTAATG

 AACTAGCTAATATCAATAGAAGGATGTACATTTCCAAATTCACAAGTTGTGTTTGATATCCAAAGCTGAA

 TACATTCTGCTTTCATCTTGGTACATACAATTATTTTACAGTTCTCCCAAGGGAGTTAGGCTATTCACA

 ACCACTCATTCAAAGTTGAAATTAACCATAGATGTAGATAAACTCAGAAATTTAATTCATGTTTCTTAAA

 TGGGCTACTTTGTCTTTTTTTGTTATTAGGGTGGTATTTAGTCTATTAGCCACAAAATTGGGAAAGGAGTAG

 AAAAAGCAGTAACTGACAACCTGAATAATACACCAGAGATAATATGAGAATCAGATCATTTCAAAACTCAT

 TTCCTATGTAAGTGCATTGAGAACTGCATATGTTTCGCTGATATATGTGTTTTTCACATTTGCGAATGGTT

Figure 2C

CCATTCTCTCCTGTACTTTTTCCAGACACTTTTTTGAGTGGATGATGTTTCGTGAAGTATACTGTATTT
TTACCTTTTTTCCTTATCACTGACACAAAAAGTAGATTAAGAGATGGGTTTGACAAGGTTCTTCCCTT
TTACATACTGCTGTCTATGTGGCTGTATCTTGTTTTCCACTACTGCTACCACAACATATTATCATGCAA
ATGCTGTATTCTTCTTTGGTGGAGATAAAGATTCTTGAGTTTGTTTTAAATTAAGCTAAAGTATCTG
TATTGCATTAAATATAATATCGACACAGTGCTTTCCGTGGCACTGCATACAATCTGAGGCCTCCTCTCTCA
GTTTTATATAGATGGCGAGAACCTAAGTTTCAGTTGATTTTACAATTGAAATGACTAAAAACAAGAAG
ACAACATTAAAAACAATATTGTTTCTA

Figure 2D

ATTAAACCTCTCGCCGAGCCCTCCGCAGACTCTGCGCCGAAAGTTTCATTTGCTGTATGCCATCCTCGA

GAGCTGTCTAGGTTAACGTTTCGCACTCTGTGTATATAACCTCGACAGTCTTGGCACCTAACGTGCTGTGCG

Met Ser Gln Trp
TAGCTGCTCCTTTGGTTGAATCCCCAGGCCCTTGTGGGGCACAAGGTGGCAGG ATG TCT CAG TGG

Tyr Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu Glu Gln Val His Gln Leu Tyr
TAC GAA CTT CAG CAG CTT GAC TCA AAA TTC CTG GAG CAG GTT CAC CAG CTT TAT

Asp Asp Ser Phe Pro Met Glu Ile Arg Gln Tyr Leu Ala Gln Trp Leu Glu Lys
GAT GAC AGT TTT CCC ATG GAA ATC AGA CAG TAC CTG GCA CAG TGG TTA GAA AAG

Gln Asp Trp Glu His Ala Ala Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His
CAA GAC TGG GAG CAC GCT GCC AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT

Asp Leu Leu Ser Gln Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn
GAC CTC CTG TCA CAG CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT AAC

Phe Leu Leu Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe
TTC TTG CTA CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG GAT AAT TTT

Gln Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys Glu Glu
CAG GAA GAC CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC TGT CTG AAG GAA GAA

Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Ser Gly Asn Ile
AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT CAG GCT CAG TCG GGG AAT ATT

Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp Ser Lys Val Arg Asn
CAG AGC ACA GTG ATG TTA GAC AAA CAG AAA GAG CTT GAC AGT AAA GTC AGA AAT

Val Lys Asp Lys Val Met Cys Ile Glu His Glu Ile Lys Ser Leu Glu Asp Leu
GTG AAG GAC AAG GTT ATG TGT ATA GAG CAT GAA ATC AAG AGC CTG GAA GAT TTA

Gln Asp Glu Tyr Asp Phe Lys Cys Lys Thr Leu Gln Asn Arg Glu His Glu Thr
CAA GAT GAA TAT GAC TTC AAA TGC AAA ACC TTG CAG AAC AGA GAA CAC GAG ACC

Asn Gly Val Ala Lys Ser Asp Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met
AAT GGT GTG GCA AAG AGT GAT CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG

Tyr Leu Met Leu Asp Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu
TAT TTA ATG CTT GAC AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG TTG

Leu Asn Val Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu
CTG AAT GTC ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA CTA GTG GAG

Trp Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
TGG AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG CCC AAT GCT TGC TTG

Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln Gln Val Arg
GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG GAG AGT CTG CAG CAA GTT CGG

Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr Thr Tyr Glu His Asp
CAG CAG CTT AAA AAG TTG GAG GAA TTG GAA CAG AAA TAC ACC TAC GAA CAT GAC

Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp Arg Thr Phe Ser Leu Phe Gln
CCT ATC ACA AAA AAC AAA CAA GTG TTA TGG GAC CGC ACC TTC AGT CTT TTC CAG

Figure 3A

Gln Leu Ile Gln Ser Ser Phe Val Val Glu Arg Gln Pro Cys Met Pro Thr His
 CAG CTC ATT CAG AGC TCG TTT GTG GTG GAA AGA CAG CCC TGC ATG CCA ACG CAC

 Pro Gln Arg Pro Leu Val Leu Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg
 CCT CAG AGG CCG CTG GTC TTG AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA

 Leu Leu Val Lys Leu Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe
 CTG TTG GTG AAA TTG CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA TTT

 Asp Lys Asp Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile
 GAT AAA GAT GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG TTC AAC ATT

 Leu Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser Leu
 TTG GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC ACC AAT GGC AGT CTG

 Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala Gly Thr Arg
 GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA CAG AAA AAT GCT GGC ACC AGA

 Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His Ser Leu Ser Phe Glu
 ACG AAT GAG GGT CCT CTC ATC GTT ACT GAA GAG CTT CAC TCC CTT AGT TTT GAA

 Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp Leu Glu Thr Thr Ser Leu Pro
 ACC CAA TTG TGC CAG CCT GGT TTG GTA ATT GAC CTC GAG ACG ACC TCT CTG CCC

 Val Val Val Ile Ser Asn Val Ser Gln Leu Pro Ser Gly Trp Ala Ser Ile Leu
 GTT GTG GTG ATC TCC AAC GTC AGC CAG CTC CCG AGC GGT TGG GCC TCC ATC CTT

 Trp Tyr Asn Met Leu Val Ala Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro
 TGG TAC AAC ATG CTG GTG GCG GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA

 Pro Cys Ala Arg Trp Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser
 CCA TGT GCA CGA TGG GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT TCT

 Val Thr Lys Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly Glu Lys Leu
 GTC ACC AAA AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA GAG AAG CTT

 Leu Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe Cys Lys
 CTT GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG ACG AGG TTT TGT AAG

 Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp Leu Trp Ile Glu Ser Ile Leu
 GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG CTT TGG ATT GAA AGC ATC CTA

 Glu Leu Ile Lys Lys His Leu Leu Pro Leu Trp Asn Asp Gly Cys Ile Met Gly
 GAA CTC ATT AAA AAA CAC CTG CTC CCT CTC TGG AAT GAT GGG TGC ATC ATG GGC

 Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu Leu Lys Asp Gln Gln Pro Gly Thr
 TTC ATC AGC AAG GAG CGA GAG CGT GCC CTG TTG AAG GAC CAG CAG CCG GGG ACC

 Phe Leu Leu Arg Phe Ser Glu Ser Ser Arg Glu Gly Ala Ile Thr Phe Thr Trp
 TTC CTG CTG CGG TTC AGT GAG AGC TCC CGG GAA GGG GCC ATC ACA TTC ACA TGG

 Val Glu Arg Ser Gln Asn Gly Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr
 GTG GAG CGG TCC CAG AAC GGA GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC

 Thr Lys Lys Glu Leu Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys
 ACG AAG AAA GAA CTT TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC AAA

 Val Met Ala Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn
 GTC ATG GCT GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG TAT CCA AAT

Figure 3B

600-1-195 A

(Sheet 12 of 42)

Ile Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu Ala Pro
ATT GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG CCA AAG GAA GCA CCA

Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr Gly Tyr Ile Lys Thr Glu Leu
GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT GGA TAT ATC AAG ACT GAG TTG

Ile Ser Val Ser Glu Val
ATT TCT GTG TCT GAA GTG TAAGTGAACACAGAAGAGTGACATGTTTACAAACCTCAAGCCAGCCT

TGCTCCTGGCTGGGGCCTGTTGAAGATGCTTGTATTTTACTTTTCCATTGTAATTGCTATCGCCATCACAG

CTGAACCTGTTGAGATCCCCGTGTTACTGCCTATCAGCATTTTACTACTTTAAAAAAAAAAAAAAAAAGCCA

AAAACCAAATTTGTATTTAAGGTATATAAATTTTCCCAAACCTGATACCCCTTGAAAAAGTATAAATAAAA

TGACCAAAAGTTGAA

Figure 3C

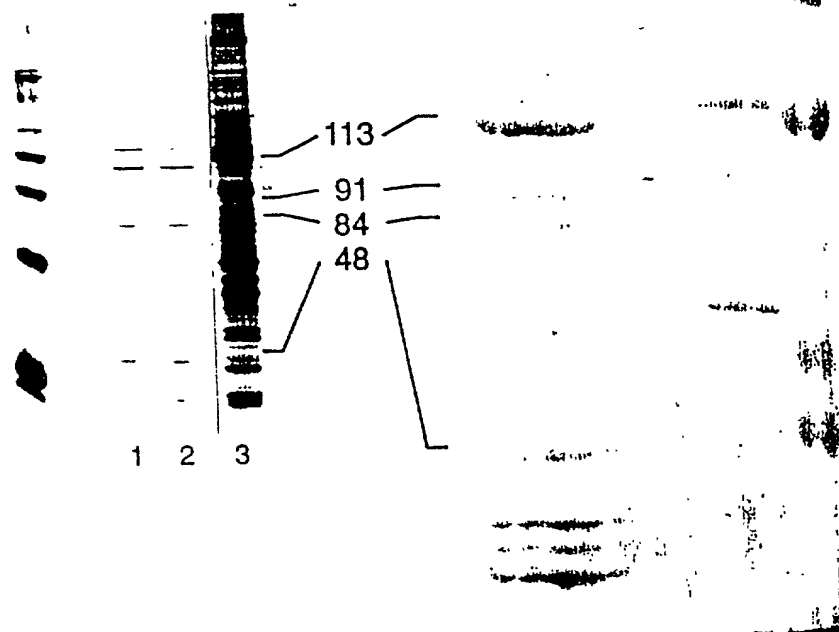


Figure 4

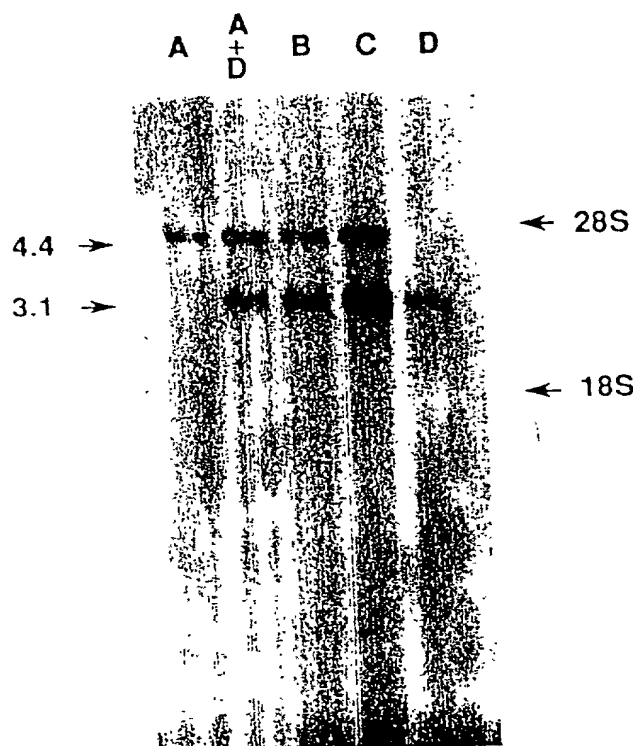
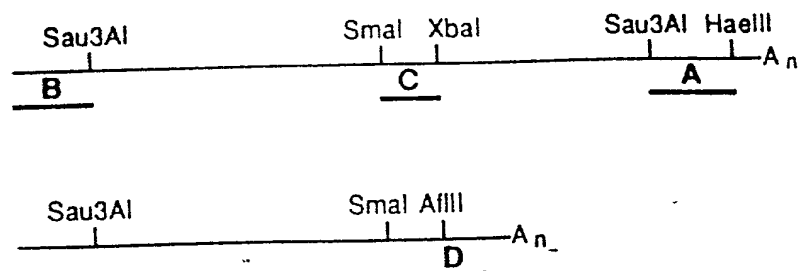


Figure 5

1 MSQWYELQQLD SKFLEQVHQLYDDSPMEIRQYLAQWLEKQDWEHAANDV
51 SFATIRFHDLDSQLDDQYSRFSLENNFLLQHINIRKSKRNLDNFQEDPIQ
101 MSMTIYSCLKEERKILENAQRFNQAQSGNIQSTVMIDKQKELDSKVRNVK
151 DKVMCIEHEIKSLEDLQDEYDFKCKTLQNRHETNGVAKSDQKQEQLLLK
201 KHYLMLDNKRKEVVHKI IELLNVTETQNALINDELVEWKRRQOSACIGG
251 PPNACLDQLQQVRQQLKKLEELEQKYTYENDPITKNKQVLWDRFSLFQQ
301 LIQSSFVVERQPCMPHTHPQRPLVLKTGVQFTVKLRLLVKLQELNYNLKVK
351 VLFDKDVNERNTVKGFRKFNILGTHKVMMEESTNGSLAAEFRHLQLKE
401 QKNAGTRTHEGPLIVTEELHLSLSFETQLCQFGLVIDLETTSLPVVISNV
451 SQLPSGWASILWYNMLVAEPRNLSFFLTPPCARWAQLSEVLSWQFSSVTK
127
501 RGLNVDOLNMLGEKLLGPNASPDGLIPWTRFCKENINDKNHFFWLWIESI
119
551 LELIKKHLLPLWNDGCGIMGFISKERERALLKDQQPGTFLLRFSESSREGA
601 ITFTWVERSONGGEPDFHAEVYTKKELSAVTFPDIIRNYKVMMAENIPE
113a
651 NPLKYLPHIDKDHAFGKYYSRPKEAPEPMELDGPKGTGYIKTELISVSE
113b
701 VHPSRLQTTDNLLPMSPEEFDEVSRIVGSVEFDSMMNTV
↑
last amino acid of 84 kd

Figure 6

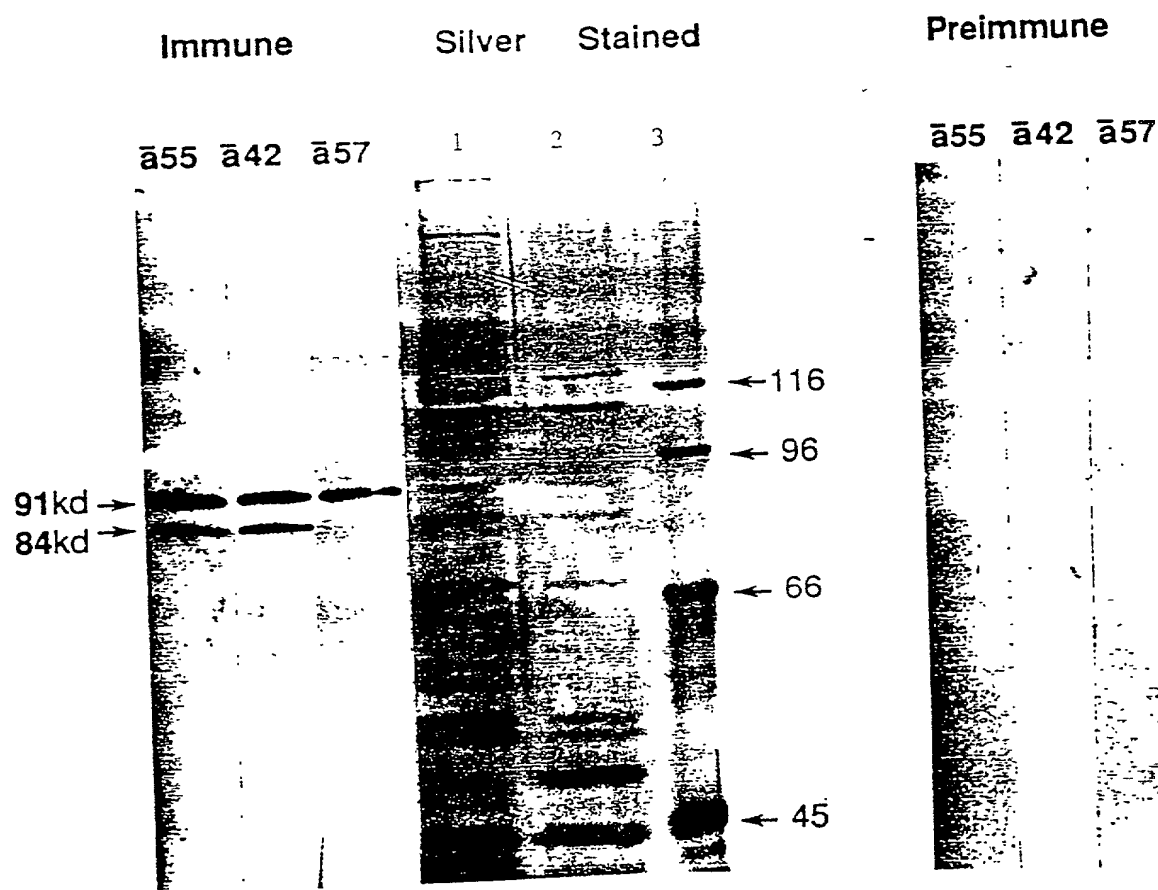


Figure 7A

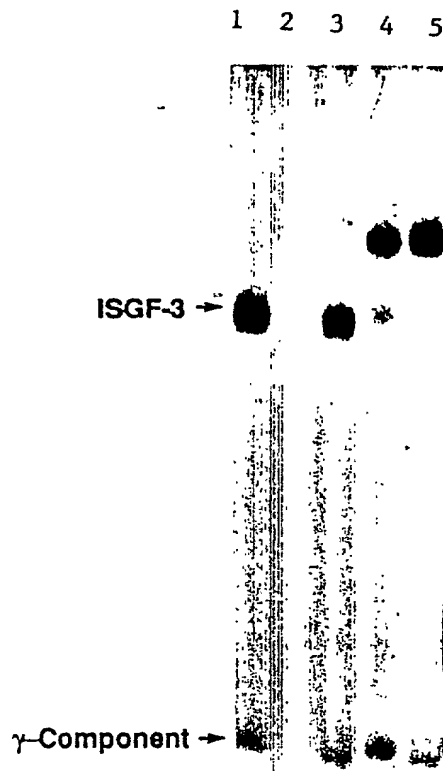


Figure 7B

1: MAQWEMLQNLDSPPQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAAALGSDDSKATMLF
61: FHFLDQLNYECGRCSQDPESLLLQHNLRKFCDRIQPFSSQDPTQLAEMIFNLLLEEKRILI
121: QAQRAQLEQGEPVLET PVESQQHEIESRILDLRAMMEKLVKSISQ LKDQQDVFCFRYKIQ
181: AKGKTPSLDPH Q TKEQKILQETLNELDKRRKEVLDASKALLGRLTTLIELLLLKLEEWKA
241: QQQKACIRAPIDHGLE EQLETWFTAGAKLLEHLRQLLKEKGLS CLVSYQDDPLTKGVDLR
301: NAQVTELLQRLLRHAFVVETQPCMPQTPHRPLILKTGSKFTVTRLLVRLQEGNESLTV
361: VSIDRNPPQLQGFRKFNIILTSNQKTLTPEKQSQGLIWDFGYLTLVEQRSGGSGKGSNKG
421: PLGVTEELHIISFTVXYTYQGLKQELKTDTLPPVVIISNMNQLSIAWASVLWFNLLSPNLQ
481: NQQFFSNPPKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLFQGNCRTEDPLLSWAD
541: FTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLKKTMSGTELL
601: RFSESSEGGITCSWVEHQDDDKVLIYSVQPYTKEVLSLPLTEIIRHYOLLTEENIPENP
661: LRFLYPRIPRDEAFGCYYQEKVNLQERRKYLKHRLIVVSNRQV DE LE Q P LE L K P P E L S
721: L S L E L G L V P E P E L S L D L E P L L K A G L D L G P S L E S V L E S T L E P V I E P T L C M V S Q T V P E P Q
781: P V S Q R V P E P D L P C D L R H L N T E P M E I F R N C V K I E E I M P N G D E L L A G Q N T V D E V V S R P S H F
841: YTDGRLMPSEF

Figure 8 A

113 kDa MAQWEHLQNLDSPEQDQLHQLYSHSLLEVDIRQYLAVWIEDQNMQEAALGSDDSKATMLF
 91/84 kDa MSQWYELQQLDSKLELGVHQLYDSDS-FMEIROMIAQMLEKQDMEHAN--NVEVFATIRE

 61 FHFLOQINYECCGRCSQDPESLLQHNLRKFCRDIQF-FSQDRTQLAEMIFNLLLEEKRII
 57 HOLUSQDDQYSRFSLE-NNFLLQHNIRKSKRNLDGNEQEDS IQMSMIYSCLKEERKII

 120 IQAQAQLEGGEPVLETPVESQCHEIESRIILDRAMHEKLVKSIISQLKQDDVFCFRYK-
 117 ENAQRFNQACSGNIQSTVMLDKQKELDSKVRNVKDKVMCIEHEIKSLEDIQDEYDEKCKT

 179 IQAKGKTPS--LOPHQTKKQKIQETLNEIKRRKEILDASKALLGRITTTIE--LLPK
 177 IQNREHETNGVAKSDQKQECLLKKMYLMQNKRRKEVVKIIEIL-NVTEITQNALINDE

 235 ILEWKAQQQKACIRARIDHGLEQLTETWFTAGAKLLFHLRQLKELKGLSCLVSYQDDHIT
 236 ILEWKKRRQQSACIGGEPNACLDQLQ-----QVROQLKMLEELEQKYYTYEHDHIT

 295 KGVDLRNAQVTELLQRILLHRNFVETQPCMFQTPHRPLILKTGSKFTVTRLLVRLQEGN
 285 KKKQVLWDRTFSLFQQLIQSSEVVERQPCMFTHPQRPLVLKTGVQFTVKIRLLVVKLQELN

 355 ESITVEMSIQRNPQP--LQGRKFNIILTSNQRTLTPEKQSQGLIWDGFIITLVEQRSG
 345 YNKKVKKLFDKDVNERNTVKGRKFNILGTHTKVHNMEESTNGSLAAEFRIQLKEQKNA

 412 GSGKGSNKGFLGVTEELHIIISFTVKYTYQGKQELKTDLPVVIIISNMNQLSIANASVLW
 405 GT--RTNEGFLIVTEELHISLSFETQLCQPGIVIDLETTSLPVVVISNVSQIPSGWASILW

 472 FNLISPNLQNOQFFSNRPKAPMSLIGPALSQWQSSYVGRGLNSDQSLMRNKIFGQNCRT
 463 YNMLVAEPRNLSFFLTPECARMAQUSEVLSWQSSVTKRGLNVQQLNMLGEKILLGPNASP

 532 EDPILSWADFTKRESPPGXIPFWTLDKILELVHDLKDLWNDGRIMGFVSRSQERRLLK
 523 DG-LIPMTRECKENINOKNFPFMLWIESILELKKHLLPLWNDGCI MGFI SKERERALLK

 592 KTHSGTFLLRFSSESS-EGGITCSHVEH-QDDDKVLIYSVQPYTKREVLSLPLTEIIRHKK
 582 DQPGTFLLRFSSESSREGAIFTWVERSCNGGEPDFHAFEPYTKKEISAVTFFDIIRNKK

 650 LLTEENIPENDIRFLYPRIPRQEAFCYY-----QEKVNLQERR--KYLKHRILVVSNR
 642 VHAENIPENDIKYLYENIDKDHAFKYYSRPEAPEPHELDGPKGTGMIKTELISVSEV

 702 QVDELQOPLELKP
 702 HPSRLQTTDNLLP

Figure 8B

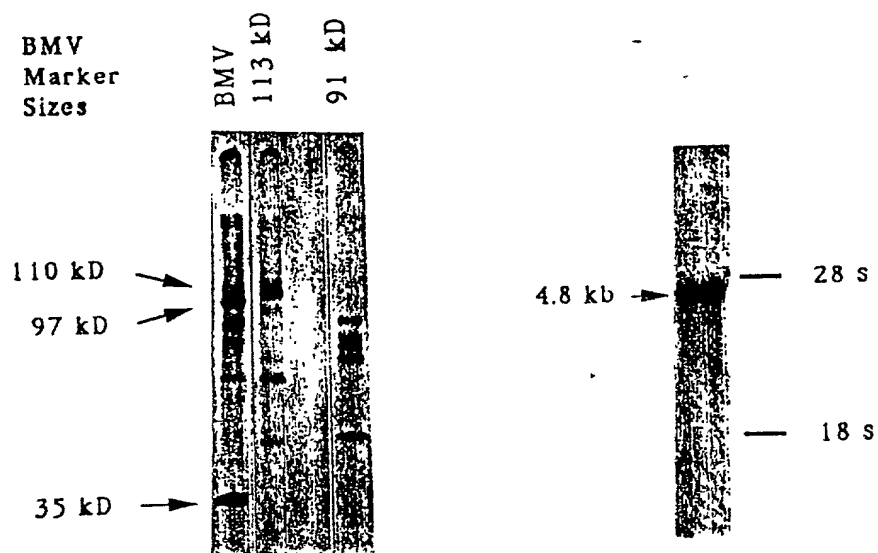


Figure 9

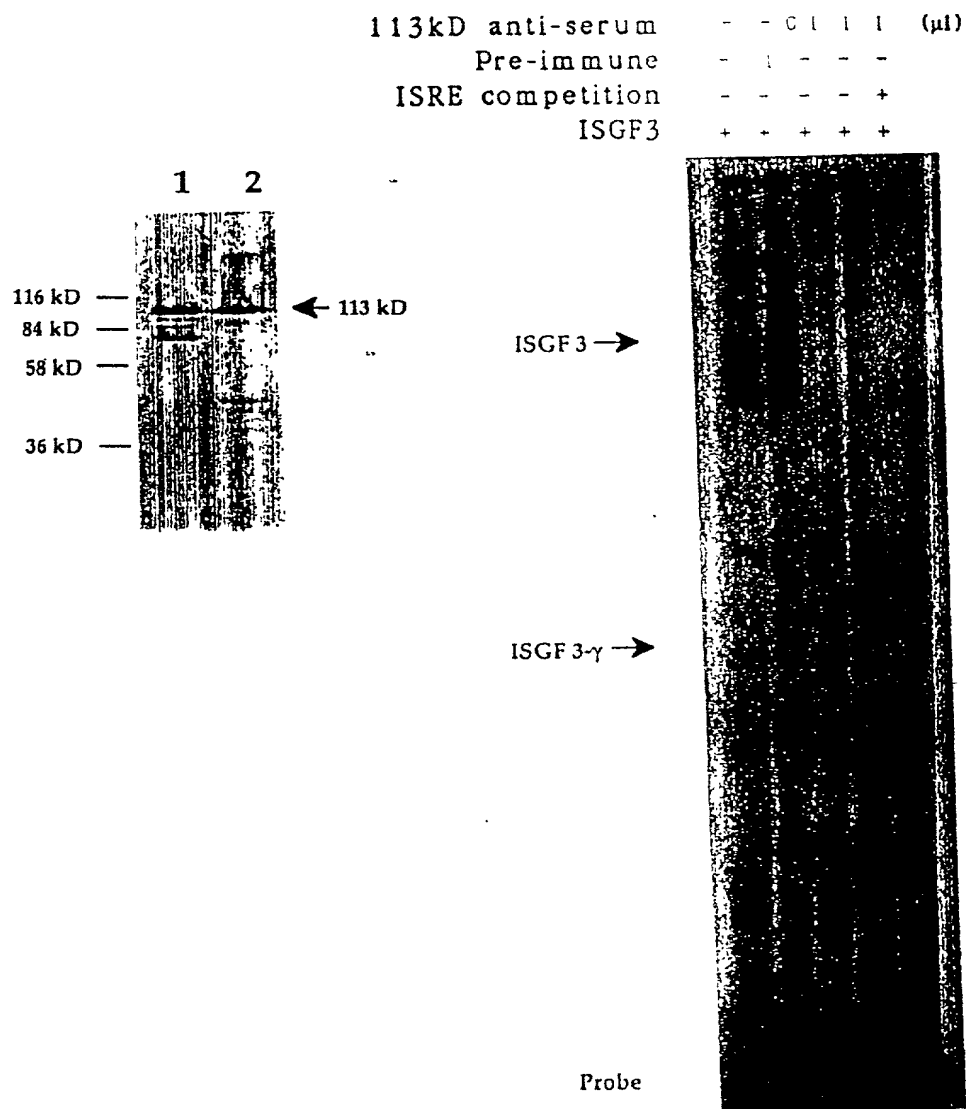


Figure 10

Figure 11

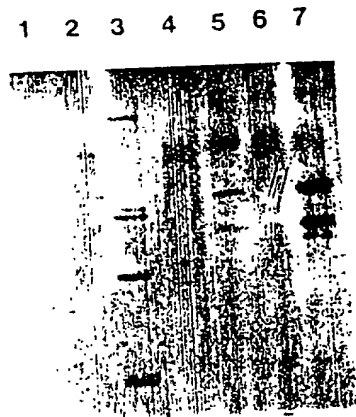


Figure 12

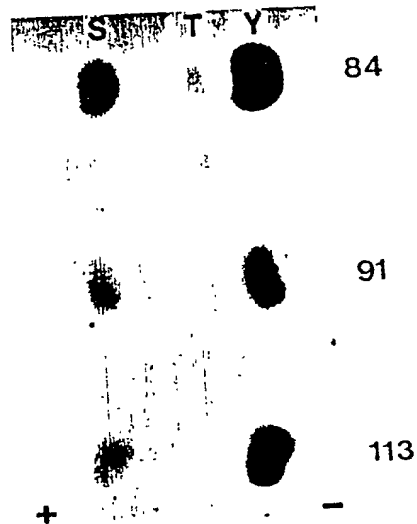


Figure 11, 12

1 MSQWFELQQL DSKFLEQVHQ LYDDSFPMEI RQYLAQWLEK QDWEHAAYDV
51 SFATIRFHD L SQLDDQYSR FSLENNFLLQ HNIRKSKRNL QDNFQEDPVQ
101 MSMIIYNCLK EERKILENAQ RFNQAQEGNI QNTVMLDKQK ELDSKVRNVK
151 DQVMCIEQEI KTLLELQDEY DFKCKTSQNR EGEANGVAKS DQKQEQLLLH
201 KMFIMLDNKR KEIIHKIREL LNSIELTQNT LINDELVEWK RRQQSACIGG
251 PPNACLDQLQ TWFTIVAETL QQIROQLKKL EELEQKFTYE PDPITKNKQV
301 LSDRTFLLFQ QLIQSSFVVE RQPCMPTHPQ RPLVLKTGVQ FTVKSRLLVK
351 LQESNLLTKV KCHFDDVNE KNTVKGFRKF NILGTHTKVM NMEESTNGSL
401 AAELRHLQLK EQKNAGNRTN EGPLIVTEEL HSLSFETQLC QPGLVIDLET
451 TSLPVVVISN VSQLP SGWAS ILWYNMLVTE PRNLSFFLNP PCAWWSQLSE
501 VLSWQFSSVT KRGLNADQLS MLGEKLLGPN AGPOGLIPWT RFCKENINDK
551 NFSFWPWIDT ILELIKNDLL CLWNDGCIMG FISKEERERL LKDQQPGTFL
601 LRFSESSREG AITFTWVERS QNGGEPDFHA VEPYTKKELS AVTFPDIIRN
651 YKVMAAENIP ENPLKYLYPN IDKDHAFGKY YSRPKEAPEP MELDDPKRTG
701 YIKTELISVS EVHPSRLQTT DNLLPMSPEE FDEMSRIVGP EFDSMMSTV

Figure 13A

1 caggatgtca cagtgggttcg agcttcagca gctggactcc aagttcctgg
51 agcaggtcca ccagctgtac gatgacagtt tccccatgga aatcagacag
101 tacctggccc agtggctgga aaagcaagac tgggagcacg ctgcctatga
151 tgtctcgttt gcgaccatcc gcttccatga cctcctctca cagctggacg
201 accagtacag ccgcttttct ctggagaata atttcttgtt gcagcacaac
251 atacggaaaa gcaagcgtaa tctccaggat aacttccaag aagatcccgt
301 acagatgtcc atgatcatct acaactgtct gaaggaagaa aggaagattt
351 tggaaaatgc ccaaagatgt aatcaggccc aggagggaaa tattcagaac
401 actgtgatgt tagataaaca gaaggagctg gacagtaaag tcagaaatgt
451 gaaggatcaa gtcattgtca tagagcagga aatcaagacc ctagaagaat
501 tacaagatga atatgacttt aaatgcaaaa cctctcagaa cagagaagggt
551 gaagccaatg gtgtggcgaa gagcgaccaa aaacaggaac agctgctgct
601 ccacaagatg tttttaatgc ttgacaataa gagaaaggag ataattcaca
651 aaatcagaga gttgctgaat tccatcgagc tcaactcagaa cactctgatt
701 aatgacgagc tcgtggagtg gaagcgaagg cagcagagcg cctgcatcgg
751 gggaccgccc aacgcctgcc tggatcagct gcaaacgtgg ttcaccattg
801 ttgcagagac cctgcagcag atccgtcagc agcttaaaaa gctggaggag
851 ttggaacaga aattcaccta tgagcccgac cctattacaa aaaacaagca
901 ggtgttgtca gatcgaacct tcctcctctt ccagcagctc attcagagct
951 ccttcgtggt agaacgacag ccgtgcatgc ccaactaccc gcagaggccc
1001 ctggctctga agactggggt acagttcact gtcaagtcca gactgttggt
1051 gaaattgcaa gagtcgaatc tattaacgaa agtgaaatgt cactttgaca
1101 aagatgtgaa cgagaaaaac acagttaaag gatttcggaa gttcaacatc
1151 ttgggtacgc acacaaaagt gatgaacatg gaagaatcca ccaacggaag
1201 tctggcagct gagctccgac acctgcaact gaaggaacag aaaaacgctg
1251 ggaacagaac taatgagggg cctctcattg tcaccgaaga acttcactct
1301 cttagctttg aaaccagtt gtgccagcca ggcttggtga ttgacctgga
1351 gaccacctct cttcctgtcg tggatgctc caacgtcagc cagctcccca

Figure 13B

1401 gtggctgggc gtctatcctg tggtaacaaca tgctgggtgac agagcccagg
1451 aatctctcct tcttcctgaa cccccctgac gcgtgggtgtt cccagctctc
1501 agagggtgtt agttggcagt tttcatcagt caccaagaga ggtctgaacg
1551 cagaccagct gagcatgctg ggagagaagc tgctgggccc taatgctggc
1601 cctgatggtc ttattccatg gacaagggtt tgtaaggaaa atattaatga
1651 taaaaatttc tccttctggc cttggattga caccatccta gagctcatta
1701 agaacgacct gctgtgcctc tgggaatgat ggtgcattat gggcttcate
1751 agcaaggagc gagaacgcgc tctgctcaag gaccagcagc cagggacgtt
1801 cctgctttaga ttcagtgaga gctcccgga aggggccatc acattcacat
1851 ggggtggaac gtcccagaac ggaggtgaac ctgacttcca tgccgtggag
1901 ccctacacga aaaaagaact ttcagctgtt actttccag atattattcg
1951 caactacaaa gtcattggctg ccgagaacat accagagaat cccctgaagt
2001 atctgtacct caatattgac aaagaccacg cctttgggaa gtattattcc
2051 agaccaaagg aagcaccaga accgatggag cttgacgacc ctaagcgaac
2101 tggatacatc aagactgagt tgatttctgt gtctgaagtc cacccttcta
2151 gacttcagac cacagacaac ctgcttccca tgtctccaga ggagtttgat
2201 gagatgtccc ggatagtggg ccccgaaatt gacagtatga tgagcacagt
2251 ataaacacga atttctctct ggagaca

Figure 13C

1 MSQWNQVQQL EIKFLEQVDQ FYDDNFPMEI RHLLAQWIET QDWEVASNNE
51 TMATILLQNL LIQLDEQLGR VSKEKNLLLI HNLKRIRKVL QGKFHGNPMH
101 VAVVISNCLR EERRILAAAN MPIQGPLEKS LQSSSVSERQ RNVEHKVSAI
151 KNSVQMTEQD TKYLEDLQDE FDYRYKTIQT MDQGDKNSIL VNQEVLTLLQ
201 EMLNSLDFKR KEALSKMTQI VNETDLLMNS MLEELQDWK KRHRIACIGG
251 PLHNGLDQLQ NCF'TLLAESL FQLRQOLEKL QEQSTKMTYE GDPIPAQRAH
301 LLERATFLIY NLFKNSFVVE RHACMP'HPQ RPMVLKTLIQ FTVKLRLLIK
351 LPELNYQVKV KASIDKNVST LSNRRFVLCG THVKAMSSEE SSMGSLSVEL
401 DIATQGDEVQ YWSKGNEGCH MVTEELHSIT FETQICLYGL TINLETSSLP
451 VVMISNVSQL PNAWASIIWY NVSTNDSQNL VFFNNPPSVT LGQLLEVMSW
501 QFSSYVGRGL NSEQLNMLAE KLTVQSNYND GHLTWAKFCK EHLPGKTFTF
551 WTWLEAILD L IKKHILPLWI DGYIMGFVSK EKERLLLKDK M'PGTFLLRFS
601 ESHLGGITFT WVDQSENGEV RFHSVEPYNK GRLSALAFAD ILRDYKVIMA
651 ENIPENPLKY LYPDIPKDKA FGKHYSSQPC EVSRPTERGD KGYVPSVFIP
701 ISTIRSDSTE PQSPSOLLPM SPSAYAVLRE NLSPTTIETA MNSPYSAE

Figure 14A

1 tgccactacc tggacggaga gagagagagc agcatgtctc agtggaatca
51 agtccaacaa ttagaaatca agtttttgga gcaagtagat cagttctatg
101 atgacaactt tcctatggaa atccggcatc tgctagctca gtggattgag
151 actcaagact gggaagtagc ttctaacaat gaaactatgg caacaattct
201 gcttcaaaac ttactaafac aattggatga acagttg^gggg cgggtttcca
251 aagaaaaaaa tctgctattg attcacaatc taaagagaat tagaaaagtt
301 cttcagggca agtttcatgg aaatccaatg catgtagctg tggtaatttc
351 aaattgctta agggaagaga ggagaatatt ggctgcagcc aacatgccta
401 tccagggacc tctggagaaa tccttacaga gttcttcagt ttctgaaaga
451 caaaggaatg tggaacacaa agtgtctgcc attaaaaaca gtgtgcagat
501 gacagaacaa gataccaaat acttagaaga cctgcaagat gagtttgact
551 acaggtataa aacaattcag acaatggatc agggtgacaa aaacagtatc
601 ctggtgaacc aggaagtttt gacactgctg caagaaatgc ttaatagtct
651 ggacttcaag agaaaggaag cactcagtaa gatgacgcag atagtgaacg
701 agacagacct gctcatgaac agcatgcttc tagaagagct gcaggactgg
751 aaaaagcggc acaggattgc ctgcattggt ggcccgcctc acaatgggct
801 ggaccagctt cagaactgct ttaccctact ggcagagagt cttttccaac
851 tcagacagca actggagaaa ctacaggagc aatctactaa aatgacctat

Figure 14B

901 gaaggggatc ccacccctgc tcaaagagca cacctcctgg aaagagctac
951 cttcctgatc tacaaccttt tcaagaactc atttgtgggc gagcgacacg
1001 catgcatgcc aacgcaccct cagaggccga tggacttaa aacctcatt
1051 cagttcactg taaaactgag attactaata aaattgccgg aactaaacta
1101 tcaggtgaaa gtaaaggcgt ccattgacaa gaatgtttca actctaagca
1151 atagaagatt tgtgctttgt ggaactcacg tcaaagctat gtccagtga
1201 gaatcttcca atgggagcct ctcaaggag ttagacattg caacccaagg
1251 agatgaagtg cagtactgga gtaaaggaaa cgagggctgc cacatggtga
1301 cagaggagtt gcattccata acctttgaga cccagatctg cctctatggc
1351 ctcaccatta acctagagac cagctcatta cctgtcgtga tgatttctaa
1401 tgtcagccaa ctacctaatg catgggcac catcatttgg tacaatgtat
1451 caactaacga ctcccagaac ttggttttct ttaataacct tccatctgtc
1501 actttgggcc aactcctgga agtgatgagc tggcaatttt catcctatgt
1551 cggctcgtggc cttaattcag agcagctcaa catgctggca gagaagctca
1601 cagttcagtc taactacaat gatggtcacc tcacctgggc caagttctgc
1651 aaggaacatt tgcttggaac aacatttacc ttctggactt ggcttgaagc
1701 aatattggac ctaattaaaa aacatattct tcccctctgg attgatggg
1751 acatcatggg atttgtagt aaagagaagg aacggcttct gctcaaagat
1801 aaaatgcctg ggacattttt gttaagattc agtgagagcc atcttgagg

Figure 14C

1851 gataaccttc acctgggtgg accaatctga aaatggagaa gtgagattcc
1901 actctgtaga accctacaac aaagggagac tgtcggtctt ggccttcgct
1951 gacatcctgc gagactacaa ggttatcatg gctgaaaaca tccctgaaaa
2001 ccctctgaag tacctctacc ctgacattcc caagacaaa gcctttggca
2051 aacactacag ctcccagccg tgcgaagtct caagaccaac cgaacgggga
2101 gacaagggtt acgtcccctc tgtttttatc cccatttcaa caatccgaag
2151 cgattccaag gagccacaat ctccctcaga ccttctcccc atgtctccaa
2201 gtgcatatgc tgtgctgaga gaaaacctga gcccaacgac aattgaaact
2251 gcaatgaatt ccccatattc tgctgaatga cggtgcaaac ggacacttta
2301 aagaaggaag cagatgaaac tggagagtgt tctttaccat agatcacaat
2351 ttattttcttc ggctttgtaa atacc

Figure 14D

1 MAQWNQLQQL DTRYLKQLHQ LYSDTFPMEL RQFLAPWIES QDWAYAASKE
51 SHATLVFHNH LGEIDQQYSR FLQESNVLYQ HNLRRIKQFL QSRYLEKPM
101 IARIVARCLW EESRLLQTAA TAAQGGQAN HPTAAVVTEK QQMLEQHLQD
151 VRKRVQDLEQ KMKVVENLQD DFDENYKTLK SQGDMQDLNG NNQSVTRQKM
201 QQLEQMLTAL QOMRRSIVSE LAGLLSAMEY VQKTLTDEEL ADWKRRPEIA
251 CIGPPNICL DRLENWITSL AESQLQTRQQ IKKLEELQOK VSYKGDPIVQ
301 HRPMLERIV ELFRNLMKSA FVVERQPCMP MHPDRPLVİK TGVQFTTKVR
351 LLVKFPELNY QLKIKVCIDK DSGDVAALRG SRKFNILGTN TKVMNMEESN
401 NGSLSAEFKH LTLREQRCGN GGRANCASL IVTEELHLIT FETEVYHQGL
451 KIDLETHSLP VVVISNICQM PNAWASILWY NMLTNNPKNV NFFTCKPIGT
501 WDQVAEVLW QFSSTTKRGL SIEQLTTLAE KLLGPGVNYS GCQITWAKFC
551 KENMAGKGF SFWWLDNIID LVKKYILALW NEGYIMGFIS KERERAILST
601 KPPGTFLRLF SESSKEGGVT FTWVEKDISG KTQIQSVEPY TKQQLNNMSF
651 AEIIMGYKIM DATNILVSPL VYLYPDIPKE EAFGKYCRPE SQEHPEADPG
701 SAAPYLKTF ICVTPTTCSN TIDLPMSPRT LDSLMQFGNN GEGAEP SAGG
751 QFESLTFDMD LTSECATSPM

Figure 15A

1 gccgcgacca gccaggccgg ccagtcgggc tcagcccga gacagtcgag
51 acccctgact gcagcaggat ggctcagtgg aaccagctgc agcagctgga
101 cacacgtac ctgaagcagc tgcaccagct gtacagcgac acgttcccca
151 tggagctgcg gcagttcctg gcaccttgga ttgagagtca agactgggca
201 tatgcagcca gcaaagagtc acatgccacg ttggtgttcc ataattcttt
251 gggtgaaatt gaccagcaat atagccgatt cctgcaagag tccaatgtcc
301 tctatcagca caaccttcga agaatacagc agtttctgca gagcaggtat
351 cttgagaagc caatggaaat tgcccggatc gtggcccgat gcctgtggga
401 agagtctcgc ctctccaga cggcagccac ggcagcccag caagggggcc
451 aggccaacca cccaacagcc gccgtagtga cagagaagca gcagatgttg
501 gagcagcatc ttcaggatgt ccggaagcga gtgcaggatc tagaacagaa
551 aatgaaggtg gtggagaacc tccaggacga ctttgatttc aactacaaaa
601 ccctcaagag ccaaggagac atgcaggatc tgaatggaaa caaccagtct
651 gtgaccagac agaagatgca gcagctggaa cagatgctca cagccctgga
701 ccagatgcgg agaagcattg tgagttagct ggcggggctc ttgtcagcaa
751 tggagtacgt gcagaagaca ctgactgatg aagagctggc tgactggaag
801 aggcggccag agatcgctg catcgaggc cctcccaaca tctgcctgga
851 ccgtctggaa aactggataa cttcattagc agaattctaa cttcagacct

Figure 15B

901 gccacaat taagaaactg gaggagctgc agcagaaagt gtcctacaag
951 ggcgacccta tcgtgcagca ccggcccatg ctggaggaga ggatcgtgga
1001 gctgttcaga aacttaatga agagtgcctt cgtgggtggag cggcagccct
1051 gcatgcccat gcacccggac cggcccttag tcatcaagac tgggtgtccag
1101 tttaccacga aagtcagggt gctgggtcaaa tttcctgagt tgaattatca
1151 gcttaaaatt aaagtgtgca ttgataaaga ctctggggat gttgctgccc
1201 tcagaggggtc tcggaaatth aacattctgg gcacgaacac aaaagtgatg
1251 aacatggagg agtctaaca cggcagcctg tctgcagagt tcaagcacct
1301 gacccttagg gagcagagat gtgggaatgg aggccgtgcc aattgtgatg
1351 cctccttgat cgtgactgag gagctgcacc tgatcacctt cgagactgag
1401 gtgtaccacc aaggcctcaa gattgaccta gagaccact ccttgccagt
1451 tgtgtgatc tccaacatct gtcagatgcc aatgcttgg gcatcaatcc
1501 tgtgttataa catgctgacc aataaccca agaacgtgaa cttcttcaact
1551 aagccgccaa ttggaacctg ggaccaagtg gccgaggtgc tcagctggca
1601 gttctcgtcc accaccaagc gagggtgag catcgagcag ctgacaacgc
1651 tggctgagaa gtccttaggg cctggtgtga actactcagg gtgtcagatc
1701 acatgggcta aattctgcaa agaaaacatg gctggcaagg gctctcctt
1751 ctgggtctgg ctgacaata tcatcgacct tgtgaaaaag tatatcttgg
1801 ccctttggaa tgaagggtac atcatgggtt tcatcgcaa ggagcgggag

Figure 15C

1851 cgggccatcc taagcacaaa gccccgggc accttctac tgcgttcag
1901 cgagagcagc aaagaaggag gggtcacttt cacttgggtg gaaaaggaca
1951 tcagtggcaa gaccagatc cagtctgtag agccatacac caagcagcag
2001 ctgaacaaca tgtcatttgc tgaaatcacc atgggctata agatcatgga
2051 tgcgaccaac atcctgggtg ctccacttgc ctacctctac cccgacattc
2101 ccaaggagga ggcatttggg aagtactgta ggcccagag ccaggagcac
2151 cccgaagccg acccaggtag tgctgccccg tacctgaaga ccaagttcat
2201 ctgtgtgaca ccaacgacct gcagcaatac cattgacctg ccgatgtccc
2251 cccgcacttt agattcattg atgcagtttg gaaataacgg tgaagggtgct
2301 gagccctcag caggagggca gtttgagtcg ctacagtttg acatggatct
2351 gacctcggag tgtgctacct ccccatgtg aggagctgaa accagaagct
2401 gcagagacgt gacttgagac acctgccccg tgctccacc ctaagcagcc
2451 gaaccccata tcgtctgaaa ctctaactt tgtgggtcca gatttttttt
2501 tttaatttcc tacttctgct atctttgggc aatctgggca ctttttaaaa
2551 gagagaaatg agtgagtgtg ggtgataaac tgttatgtaa agaggagaga
2601 cctctgagtc tggggatggg gctgagagca gaaggaggc aaaggggaac
2651 acctcctgtc ctgcccgcct gccctccttt ttcagcagct cgggggttgg
2701 ttgttagaca agtgccctct ggtgcccatt gctacctgtt gcccactct
2751 gtgagctgat acccattctt gggaactcct ggctctgcac tttcaacctt

Figure 15D

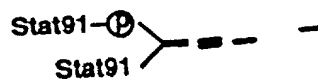
2801 gctaatatcc acatagaagc taggactaag cccaggaggt tcctctttaa

2851 attaaaaaaaa aaaaaaaaaa

Figure 15E

A

Ext
Flow
A_{0.2}
A_{0.8}



B

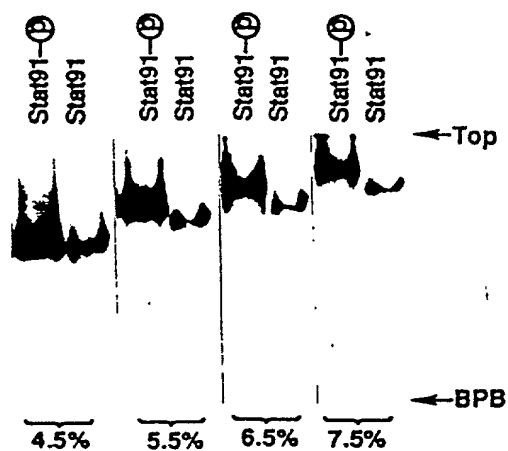


Figure 16A, B

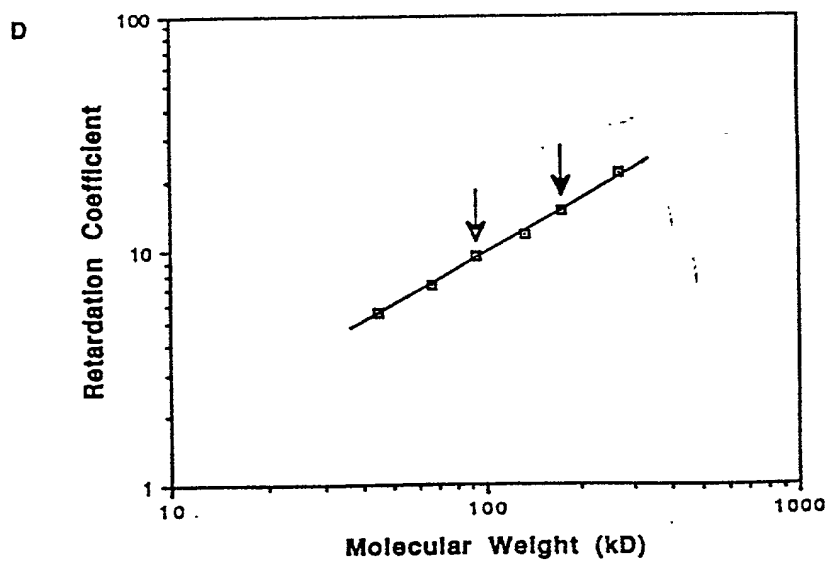
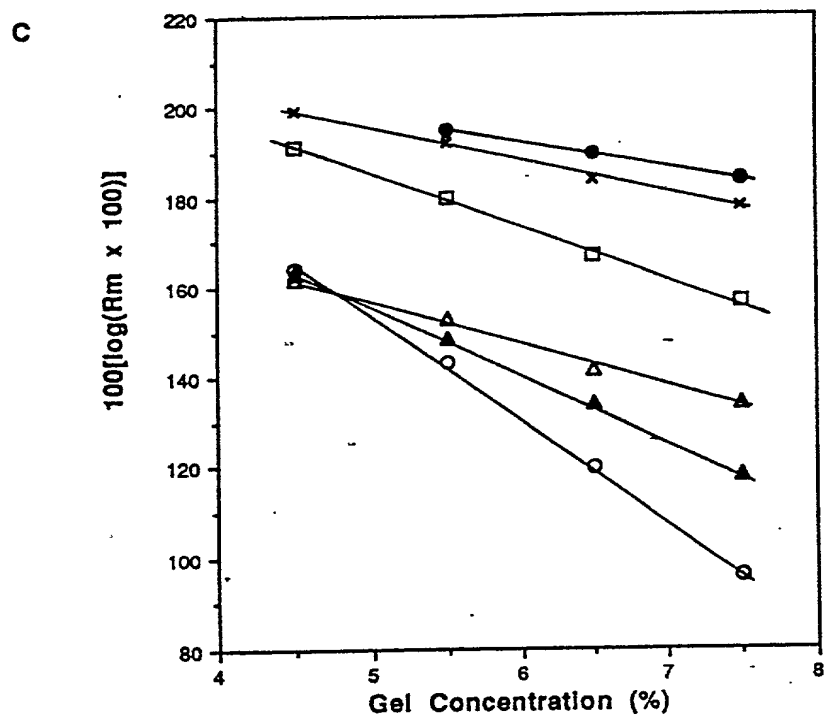


Figure 16C,D

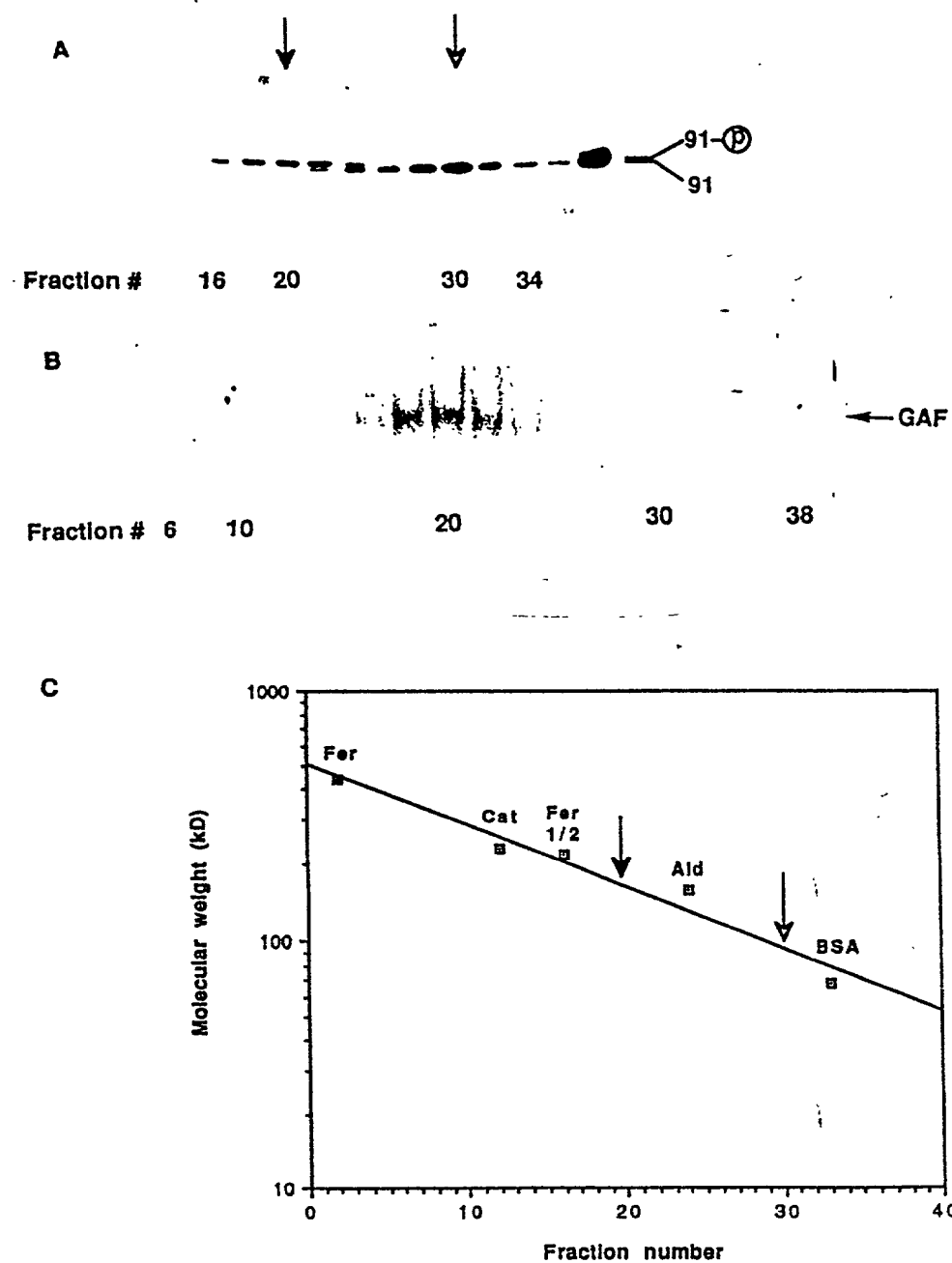


Figure 17

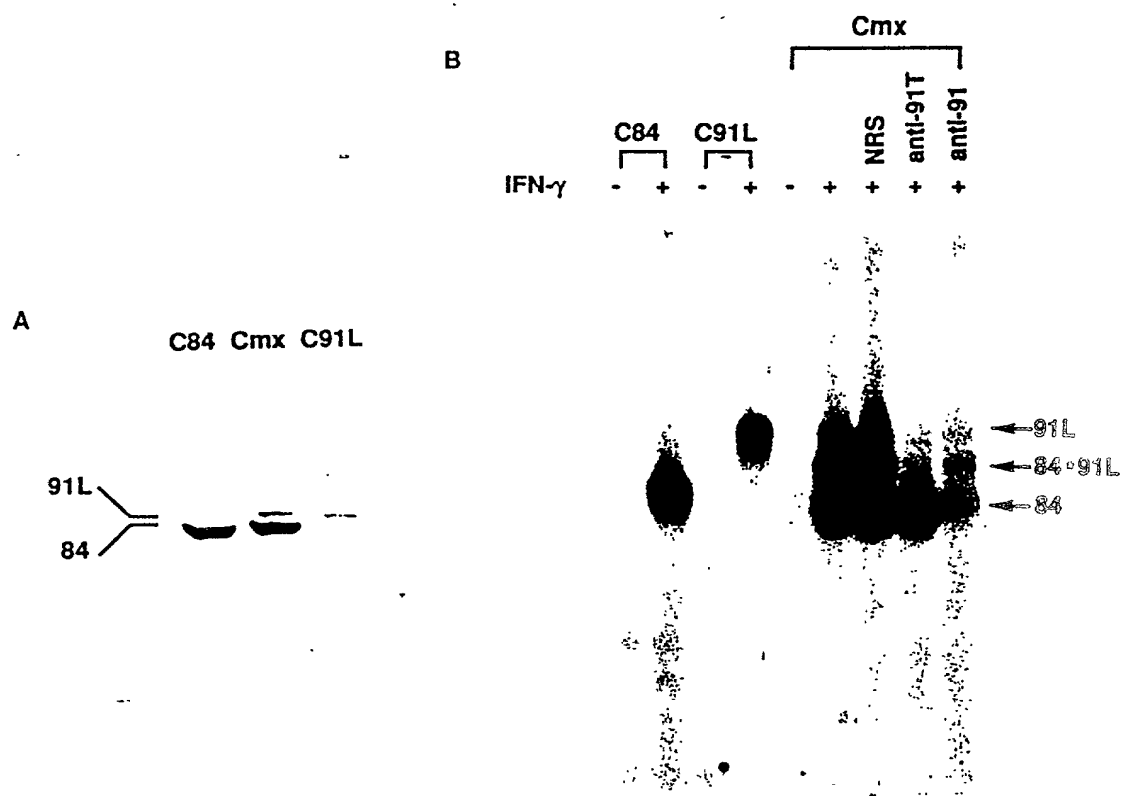


Figure 18

Figure 19

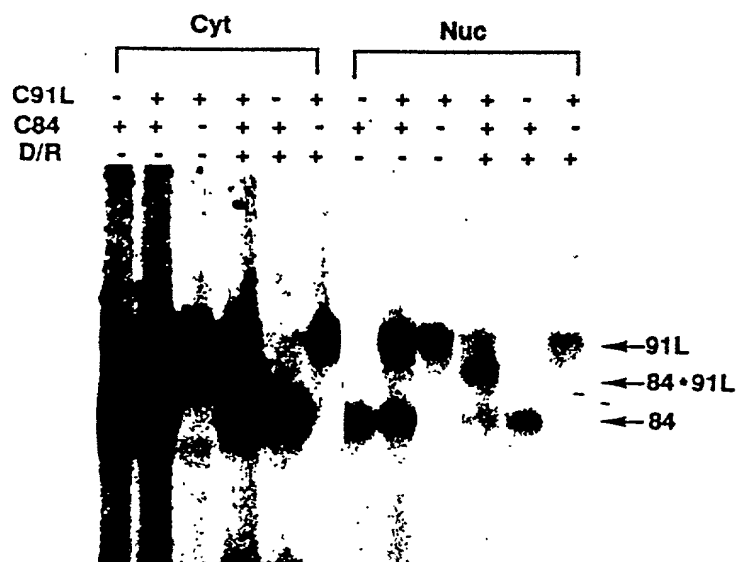


Figure 20

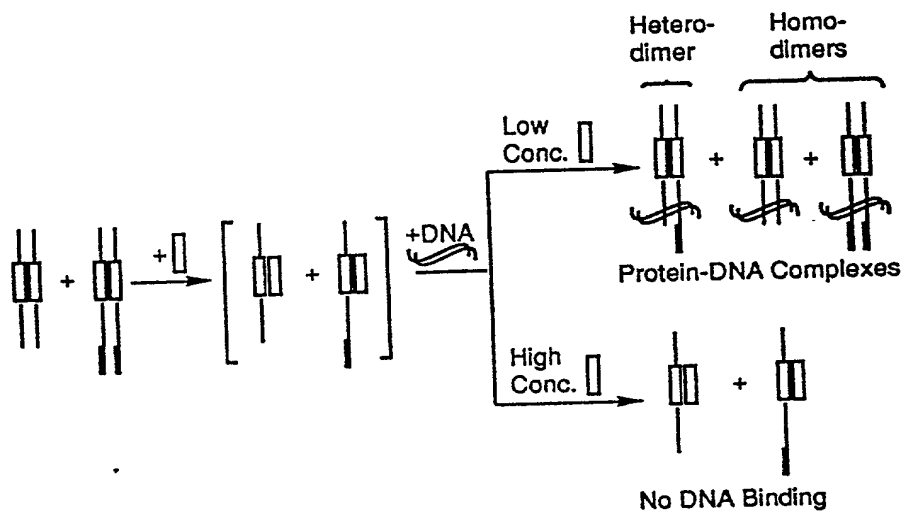


Figure 19, 20

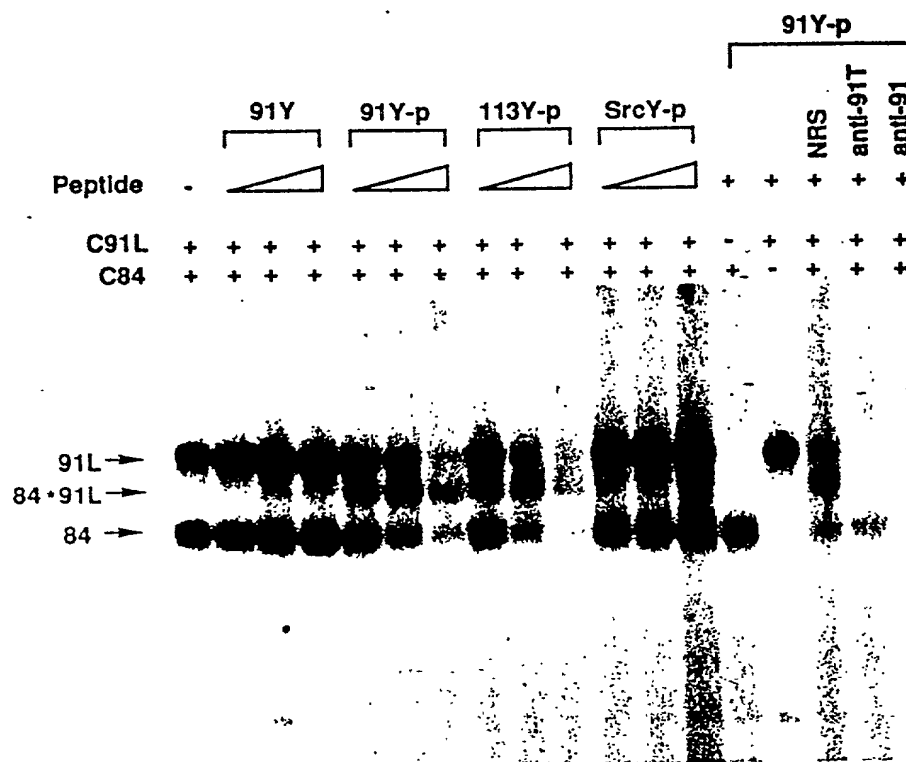


Figure 21

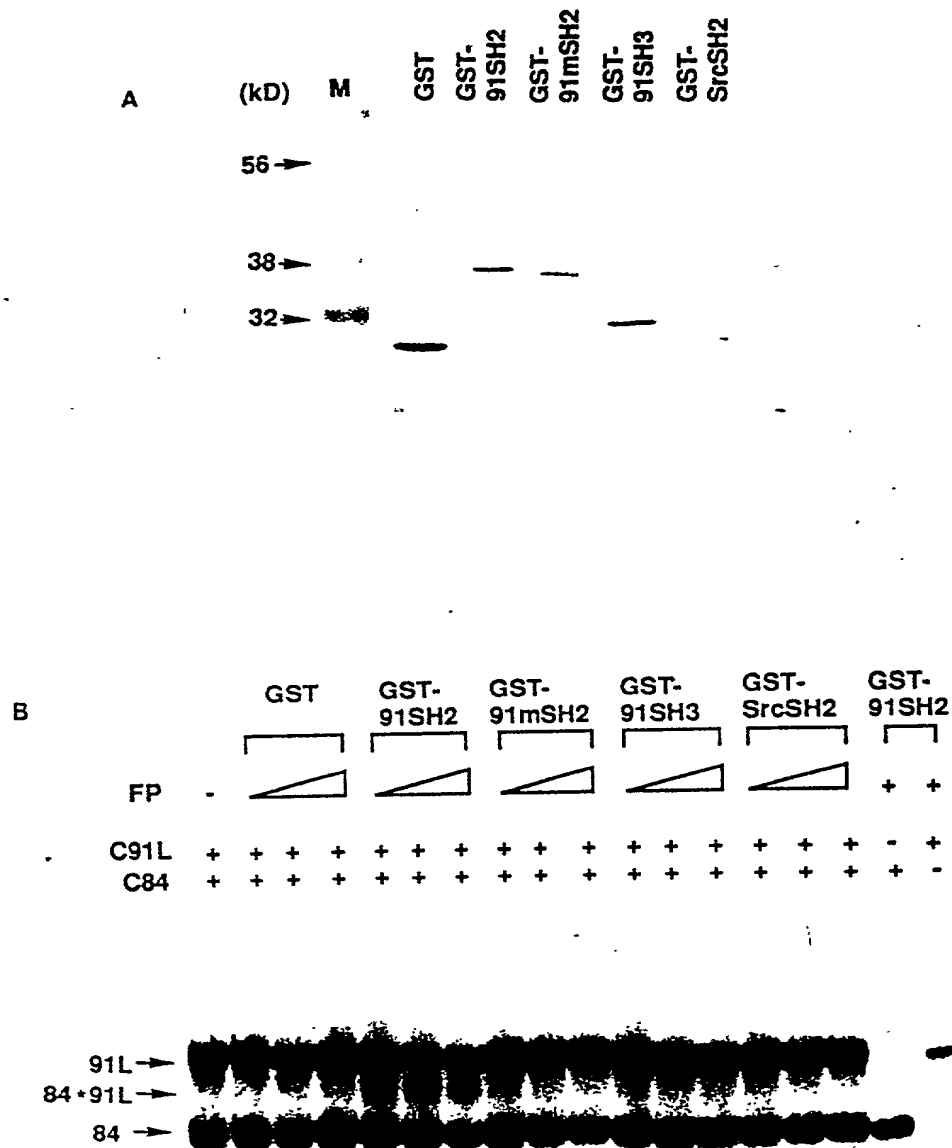


Figure 22

		$\beta A1$	$\alpha A2$	$\beta B5$	
stat91	(569)	LLPL WND GRCIMGFI	SKERERALLK DQOP	G TFLLRFS	ESSRE G AITFWVER (619)
src	(145)	AEE WYF GKI	TRRESERLLL NPENPRG	TFLVRES ETTK	G AYCLSVSD (188)
lck	(127)	WFF KNL	SRKDAERQLL APGNTHG	SFLIRES ESTA	G SFSLSVRD (168)
abl	(141)	EKHS WYH GPV	SRNAAEYLLS SGIN	G SFLVRES DRRP	G QRSISLRY (184)
p85 α N	(330)	QDAE WYW GDI	SREEVNEKLR DTAD	G TFLVRDA STKM	G DYTTLTRK (374)

SCR'S	XXX	XXXXXXXXXX	XXXXX	XXX	XXXXX
Name	NA βA	AA	αA	AB	βB BC βC

		$\beta D6$	
stat91	(620)	S QN	GGEPDFHAVEPYTKKELSAVTFP IIRNYKV MAAENIPENPL (664)
src	(189)	F FD NAK GL	NVKHYKI RKLDS G (210)
lck	(169)	D FD QNQ GE	VVKHYKI RNLDN G (189)
abl	(185)	E E G	RVYHYRI NTASD G (200)
p85 α N	(375)	GG	NNKLIKI FHR D G (388)

SCR'S	XXXXXXXX X	X
Name	CD	βD $\beta D'$ DE

		$\alpha B9$	
stat91	(665)	KYLY P NID K	KDHAFGKYYSRP PKEA PEP M ELD GPKGTGYIKT (704)
src	(211)	GFYI TSR TQF S	SLQQLVAYYSKH ADGLCH RLT NVCPTS (248)
lck	(190)	GFYI SPR ITF P	GLHDLVRHYTNA SDGLCT RLS RPCQTQ (227)
abl	(201)	KLYV SSE SRF N	TLAELVHHSTV ADGLIT TLH YPA PKR (238)
p85 α N	(389)	KYGF SDP LTF N	SVVELINHYRHE S LAQYN PKLDV KL LYP (427)

SCR'S	XXX	XXXXXXXXXX	
Name	βE EF βF	αB	BG βG GQ

Figure 23

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As below named inventors, we hereby declare that:

Our residence, post office address and citizenship are as stated below under our names.

We believe that we are the original, first and joint inventors of the subject matter which is claimed and for which a patent is sought on the invention entitled

**RECEPTOR RECOGNITION FACTORS, PROTEIN SEQUENCES
AND METHODS OF USE THEREOF**

the Specification of which was filed on March 11, 1994 and was accorded U.S. Serial No. 08/212,185.

We hereby state that we have reviewed and understand the contents of the above-identified Specification, including the claims, as amended by any amendment referred to above.

We acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, 1.56(a).

We hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

<u>APPLICATION</u>	<u>PRIOR FILED APPLICATION(S)</u>	
<u>NUMBER</u>	<u>COUNTRY (DAY/MONTH/YEAR FILED)</u>	<u>PRIORITY CLAIMED</u>
NONE		

We hereby claim the benefit under Title 35, United States Code, §120 of any United States application listed below, and, insofar as the subject matter of each of the claims of this application is not disclosed in any prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, we acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a), which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

<u>APPLICATION</u>	<u>FILING DATE</u>	<u>STATUS - PATENTED, PENDING,</u>
<u>NO.</u>	<u>(DAY/MONTH/YEAR)</u>	<u>ABANDONED</u>
08/126,588	24/9/93	PENDING
07/854,296	19/03/92	PENDING
07/980,498	23/11/92	PENDING
PCT US93/02569	19/3/93	PENDING

I hereby appoint as my attorneys or agents the following persons: Stefan J. Klauber (Attorney, Registration No. 22,604); David A. Jackson (Attorney, Registration No. 26,742); Barbara L. Renda (Attorney, Registration No. 27,626); Paul F. Fehlner (Agent, Registration No. 35,135); and Thomas E. Anderson (Agent, Registration No. 37,063), said attorneys or agents with full power of substitution and revocation to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Please address all correspondence regarding this application to:

DAVID A. JACKSON, ESQ.
KLAUBER & JACKSON
411 HACKENSACK AVENUE
HACKENSACK, NEW JERSEY 07601

Direct all telephone calls to David A. Jackson at (201) 487-5800.

We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF INVENTOR: JAMES E. DARNELL, JR.
RESIDENCE: 22 Chestnut Avenue
Larchmont, New York 10538

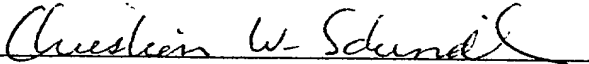
COUNTRY OF CITIZENSHIP: USA
POST OFFICE ADDRESS: Same

SIGNATURE OF INVENTOR 

DATE 5/18/94

FULL NAME OF INVENTOR: CHRISTIAN W. SCHINDLER
RESIDENCE: 18 Terrell Avenue, Rockville Center
New York, New York 11570

COUNTRY OF CITIZENSHIP: USA
POST OFFICE ADDRESS: Same

SIGNATURE OF INVENTOR 

DATE 5/26/94

FULL NAME OF INVENTOR: XIN-YUAN FU
RESIDENCE: 1249 Park Avenue, Apt. 2A 68-37 Yellowstone Blvd., Apt. C6
New York, New York 10029 Forrest Hills, N.Y. 11375
COUNTRY OF CITIZENSHIP: CHINA
POST OFFICE ADDRESS: Same
SIGNATURE OF INVENTOR Xin-yuan Fu
DATE 5/23/94

FULL NAME OF INVENTOR: ZILONG WEN
RESIDENCE: 1230 York Avenue
New York, New York 10021-6399
COUNTRY OF CITIZENSHIP: CHINA
POST OFFICE ADDRESS: Same
SIGNATURE OF INVENTOR Zilong Wen
DATE 5/23/94

FULL NAME OF INVENTOR: ZHONG ZHONG
RESIDENCE: 1230 York Avenue
New York, New York 10021-6399
COUNTRY OF CITIZENSHIP: CHINA
POST OFFICE ADDRESS: Same
SIGNATURE OF INVENTOR Zhong Zhong
DATE 5/23/94

PATENT
600-1-195A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT : JAMES E. DARNELL, JR. ET AL.
SERIAL NO. : UNASSIGNED EXAMINER : UNKNOWN
FILED : HEREWITH ART UNIT : UNKNOWN
FOR : NUCLEIC ACIDS ENCODING RECEPTOR RECOGNITION
FACTORS AND METHODS OF USE THEREOF (AMENDED)

ASSOCIATE POWER OF ATTORNEY

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

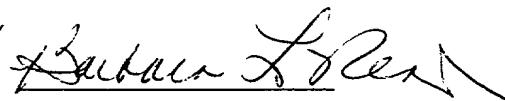
Dear Sir:

The undersigned attorney for applicant hereby appoints Michael D. Davis,
Registration No. 39,161, as Associate Attorney to prosecute the above-identified patent
application and to transact all business in the Patent and Trademark Office connected
herewith.

Please continue to address all correspondence as follows:

David A. Jackson
KLAUBER & JACKSON
411 Hackensack Avenue
Hackensack, New Jersey 07601

Respectfully submitted,



BARBARA L. RENDA
Attorney for Applicant(s)
Registration No. 27,626

KLAUBER & JACKSON
411 Hackensack Avenue
Hackensack, New Jersey 07601
(201) 487-5800

Date: October 10, 1997

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Darnell Jr., James E.
Schindler, Christian W.
Fu, Xian-Yuan
Wen, Zilong
Zhong, Zhong
- (ii) TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
SEQUENCES AND METHODS OF USE THEREOF
- (iii) NUMBER OF SEQUENCES: 25
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Klauber & Jackson
 - (B) STREET: 411 Hackensack Avenue
 - (C) CITY: Hackensack
 - (D) STATE: New Jersey
 - (E) COUNTRY: USA
 - (F) ZIP: 07601
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/212,185
 - (B) FILING DATE: 11-MAR-1994
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/980,498
 - (B) FILING DATE: 23-NOV-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/854,296
 - (B) FILING DATE: 19-MAR-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: WO US93/02569
 - (B) FILING DATE: 19-MAR-1993
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/126,588
 - (B) FILING DATE: 24-SEP-1993
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Jackson Esq., David A.
 - (B) REGISTRATION NUMBER: 26,742
 - (C) REFERENCE/DOCKET NUMBER: 600-1-073 CIP
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 201 487-5800
 - (B) TELEFAX: 201 343-1684
 - (C) TELEX: 133521

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3268 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (vii) IMMEDIATE SOURCE:
 (B) CLONE: HeLa
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 25..2577

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACTGCAACCC TAATCAGAGC CCAA ATG GCG CAG TGG GAA ATG CTG CAG AAT	51
Met Ala Gln Trp Glu Met Leu Gln Asn	
1 5	
CTT GAC AGC CCC TTT CAG GAT CAG CTG CAC CAG CTT TAC TCG CAC AGC	99
Leu Asp Ser Pro Phe Gln Asp Gln Leu His Gln Leu Tyr Ser His Ser	
10 15 20 25	
CTC CTG CCT GTG GAC ATT CGA CAG TAC TTG GCT GTC TGG ATT GAA GAC	147
Leu Leu Pro Val Asp Ile Arg Gln Tyr Leu Ala Val Trp Ile Glu Asp	
30 35 40	
CAG AAC TGG CAG GAA GCT GCA CTT GGG AGT GAT GAT TCC AAG GCT ACC	195
Gln Asn Trp Gln Glu Ala Ala Leu Gly Ser Asp Asp Ser Lys Ala Thr	
45 50 55	
ATG CTA TTC TTC CAC TTC TTG GAT CAG CTG AAC TAT GAG TGT GGC CGT	243
Met Leu Phe Phe His Phe Leu Asp Gln Leu Asn Tyr Glu Cys Gly Arg	
60 65 70	
TGC AGC CAG GAC CCA GAG TCC TTG TTG CTG CAG CAC AAT TTG CGG AAA	291
Cys Ser Gln Asp Pro Glu Ser Leu Leu Leu Gln His Asn Leu Arg Lys	
75 80 85	
TTC TGC CGG GAC ATT CAG CCC TTT TCC CAG GAT CCT ACC CAG TTG GCT	339
Phe Cys Arg Asp Ile Gln Pro Phe Ser Gln Asp Pro Thr Gln Leu Ala	
90 95 100 105	
GAG ATG ATC TTT AAC CTC CTT CTG GAA GAA AAA AGA ATT TTG ATC CAG	387
Glu Met Ile Phe Asn Leu Leu Glu Glu Lys Arg Ile Leu Ile Gln	
110 115 120	
GCT CAG AGG GCC CAA TTG GAA CAA GGA GAG CCA GTT CTC GAA ACA CCT	435
Ala Gln Arg Ala Gln Leu Glu Gln Gly Glu Pro Val Leu Glu Thr Pro	
125 130 135	
GTG GAG AGC CAG CAA CAT GAG ATT GAA TCC CGG ATC CTG GAT TTA AGG	483
Val Glu Ser Gln Gln His Glu Ile Glu Ser Arg Ile Leu Asp Leu Arg	
140 145 150	
GCT ATG ATG GAG AAG CTG GTA AAA TCC ATC AGC CAA CTG AAA GAC CAG	531
Ala Met Met Glu Lys Leu Val Lys Ser Ile Ser Gln Leu Lys Asp Gln	
155 160 165	
CAG GAT GTC TTC TGC TTC CGA TAT AAG ATC CAG GCC AAA GGG AAG ACA	579
Gln Asp Val Phe Cys Phe Arg Tyr Lys Ile Gln Ala Lys Gly Lys Thr	
170 175 180 185	

CCC	TCT	CTG	GAC	CCC	CAT	CAG	ACC	AAA	GAG	CAG	AAG	ATT	CTG	CAG	GAA	627
Pro	Ser	Leu	Asp	Pro	His	Gln	Thr	Lys	Glu	Gln	Lys	Ile	Leu	Gln	Glu	
				190					195					200		
ACT	CTC	AAT	GAA	CTG	GAC	AAA	AGG	AGA	AAG	GAG	GTG	CTG	GAT	GCC	TCC	675
Thr	Leu	Asn	Glu	Leu	Asp	Lys	Arg	Arg	Lys	Glu	Val	Leu	Asp	Ala	Ser	
			205					210					215			
AAA	GCA	CTG	CTA	GGC	CGA	TTA	ACT	ACC	CTA	ATC	GAG	CTA	CTG	CTG	CCA	723
Lys	Ala	Leu	Leu	Gly	Arg	Leu	Thr	Thr	Leu	Ile	Glu	Leu	Leu	Leu	Pro	
		220					225					230				
AAG	TTG	GAG	GAG	TGG	AAG	GCC	CAG	CAG	CAA	AAA	GCC	TGC	ATC	AGA	GCT	771
Lys	Leu	Glu	Glu	Trp	Lys	Ala	Gln	Gln	Gln	Lys	Ala	Cys	Ile	Arg	Ala	
	235					240					245					
CCC	ATT	GAC	CAC	GGG	TTG	GAA	CAG	CTG	GAG	ACA	TGG	TTC	ACA	GCT	GGA	819
Pro	Ile	Asp	His	Gly	Leu	Glu	Gln	Leu	Glu	Thr	Trp	Phe	Thr	Ala	Gly	
				255						260					265	
GCA	AAG	CTG	TTG	TTT	CAC	CTG	AGG	CAG	CTG	CTG	AAG	GAG	CTG	AAG	GGA	867
Ala	Lys	Leu	Leu	Phe	His	Leu	Arg	Gln	Leu	Leu	Lys	Glu	Leu	Lys	Gly	
				270					275					280		
CTG	AGT	TGC	CTG	GTT	AGC	TAT	CAG	GAT	GAC	CCT	CTG	ACC	AAA	GGG	GTG	915
Leu	Ser	Cys	Leu	Val	Ser	Tyr	Gln	Asp	Asp	Pro	Leu	Thr	Lys	Gly	Val	
			285					290					295			
GAC	CTA	CGC	AAC	GCC	CAG	GTC	ACA	GAG	TTG	CTA	CAG	CGT	CTG	CTC	CAC	963
Asp	Leu	Arg	Asn	Ala	Gln	Val	Thr	Glu	Leu	Leu	Gln	Arg	Leu	Leu	His	
		300					305					310				
AGA	GCC	TTT	GTG	GTA	GAA	ACC	CAG	CCC	TGC	ATG	CCC	CAA	ACT	CCC	CAT	1011
Arg	Ala	Phe	Val	Val	Glu	Thr	Gln	Pro	Cys	Met		Gln	Thr	Pro	His	
		315				320					325					
CGA	CCC	CTC	ATC	CTC	AAG	ACT	GGC	AGC	AAG	TTC	ACC	GTC	CGA	ACA	AGG	1059
Arg	Pro	Leu	Ile	Leu	Lys	Thr	Gly	Ser	Lys	Phe	Thr	Val	Arg	Thr	Arg	
					335					340					345	
CTG	CTG	GTG	AGA	CTC	CAG	GAA	GGC	AAT	GAG	TCA	CTG	ACT	GTG	GAA	GTC	1107
Leu	Leu	Val	Arg	Leu	Gln	Glu	Gly	Asn	Glu	Ser	Leu	Thr	Val	Glu	Val	
				350					355					360		
TCC	ATT	GAC	AGG	AAT	CCT	CCT	CAA	TTA	CAA	GGC	TTC	CGG	AAG	TTC	AAC	1155
Ser	Ile	Asp	Arg	Asn	Pro	Pro	Gln	Leu	Gln	Gly	Phe	Arg	Lys	Phe	Asn	
			365					370					375			
ATT	CTG	ACT	TCA	AAC	CAG	AAA	ACT	TTG	ACC	CCC	GAG	AAG	GGG	CAG	AGT	1203
Ile	Leu	Thr	Ser	Asn	Gln	Lys	Thr	Leu	Thr	Pro	Glu	Lys	Gly	Gln	Ser	
			380				385					390				
CAG	GGT	TTG	ATT	TGG	GAC	TTT	GGT	TAC	CTG	ACT	CTG	GTG	GAG	CAA	CGT	1251
Gln	Gly	Leu	Ile	Trp	Asp	Phe	Gly	Tyr	Leu	Thr	Leu	Val	Glu	Gln	Arg	
		395				400					405					
TCA	GGT	GGT	TCA	GGA	AAG	GGC	AGC	AAT	AAG	GGG	CCA	CTA	GGT	GTG	ACA	1299
Ser	Gly	Gly	Ser	Gly	Lys	Gly	Ser	Asn	Lys	Gly	Pro	Leu	Gly	Val	Thr	
					415					420					425	
GAG	GAA	CTG	CAC	ATC	ATC	AGC	TTC	ACG	GTC	AAA	TAT	ACC	TAC	CAG	GGT	1347
Glu	Glu	Leu	His	Ile	Ile	Ser	Phe	Thr	Val	Lys	Tyr	Thr	Tyr	Gln	Gly	
				430					435					440		
CTG	AAG	CAG	GAG	CTG	AAA	ACG	GAC	ACC	CTC	CCT	GTG	GTG	ATT	ATT	TCC	1395
Leu	Lys	Gln	Glu	Leu	Lys	Thr	Asp	Thr	Leu	Pro	Val	Val	Ile	Ile	Ser	
			445					450					455			

AAC Asn	ATG Met	AAC Asn 460	CAG Gln	CTC Leu	TCA Ser	ATT Ile	GCC Ala 465	TGG Trp	GCT Ala	TCA Ser	GTT Val	CTC Leu 470	TGG Trp	TTC Phe	AAT Asn	1443
TTG Leu 475	CTC Leu	AGC Ser	CCA Pro	AAC Asn	CTT Leu	CAG Gln 480	AAC Asn	CAG Gln	CAG Gln	TTC Phe	TTC Phe 485	TCC Ser	AAC Asn	CCC Pro	CCC Pro	1491
AAG Lys 490	GCC Ala	CCC Pro	TGG Trp	AGC Ser	TTG Leu 495	CTG Leu	GGC Gly	CCT Pro	GCT Ala	CTC Leu 500	AGT Ser	TGG Trp	CAG Gln	TTC Phe	TCC Ser 505	1539
TCC Ser	TAT Tyr	GTT Val	GGC Gly	CGA Arg 510	GGC Gly	CTC Leu	AAC Asn	TCA Ser	GAC Asp 515	CAG Gln	CTG Leu	AGC Ser	ATG Met	CTG Leu 520	AGA Arg	1587
AAC Asn	AAG Lys	CTG Leu 525	TTC Phe	GGG Gly	CAG Gln	AAC Asn	TGT Cys 530	AGG Arg	ACT Thr	GAG Glu	GAT Asp	CCA Pro	TTA Leu 535	TTG Leu	TCC Ser	1635
TGG Trp	GCT Ala	GAC Asp 540	TTC Phe	ACT Thr	AAG Lys	CGA Arg	GAG Glu 545	AGC Ser	CCT Pro	CCT Pro	GGC Gly	AAG Lys 550	TTA Leu	CCA Pro	TTC Phe	1683
TGG Trp 555	ACA Thr	TGG Trp	CTG Leu	GAC Asp	AAA Lys	ATT Ile 560	CTG Leu	GAG Glu	TTG Leu	GTA Val	CAT His 565	GAC Asp	CAC His	CTG Leu	AAG Lys	1731
GAT Asp 570	CTC Leu	TGG Trp	AAT Asn	GAT Asp	GGA Gly 575	CGC Arg	ATC Ile	ATG Met	GGC Gly	TTT Phe 580	GTG Val	AGT Ser	CGG Arg	AGC Ser	CAG Gln 585	1779
GAG Glu	CGC Arg	CGG Arg	CTG Leu	CTG Leu 590	AAG Lys	AAG Lys	ACC Thr	ATG Met	TCT Ser 595	GGC Gly	ACC Thr	TTT Phe	CTA Leu	CTG Leu 600	CGC Arg	1827
TTC Phe	AGT Ser	GAA Glu 605	TCG Ser	TCA Ser	GAA Glu	GGG Gly	GGC Gly	ATT Ile 610	ACC Thr	TGC Cys	TCC Ser	TGG Trp	GTG Val 615	GAG Glu	CAC His	1875
CAG Gln	GAT Asp	GAT Asp 620	GAC Asp	AAG Lys	GTG Val	CTC Leu	ATC Ile 625	TAC Tyr	TCT Ser	GTG Val	CAA Gln	CCG Pro 630	TAC Tyr	ACG Thr	AAG Lys	1923
GAG Glu 635	GTG Val	CTG Leu	CAG Gln	TCA Ser	CTC Leu	CCG Pro 640	CTG Leu	ACT Thr	GAA Glu	ATC Ile	ATC Ile 645	CGC Arg	CAT His	TAC Tyr	CAG Gln	1971
TTG Leu 650	CTC Leu	ACT Thr	GAG Glu	GAG Glu	AAT Asn 655	ATA Ile	CCT Pro	GAA Glu	AAC Asn	CCA Pro 660	CTG Leu	CGC Arg	TTC Phe	CTC Leu	TAT Tyr 665	2019
CCC Pro	CGA Arg	ATC Ile	CCC Pro	CGG Arg 670	GAT Asp	GAA Glu	GCT Ala	TTT Phe	GGG Gly 675	TGC Cys	TAC Tyr	TAC Tyr	CAG Gln	GAG Glu 680	AAA Lys	2067
GTT Val	AAT Asn	CTC Leu	CAG Gln 685	GAA Glu	CGG Arg	AGG Arg	AAA Lys	TAC Tyr 690	CTG Leu	AAA Lys	CAC His	AGG Arg	CTC Leu 695	ATT Ile	GTG Val	2115
GTC Val	TCT Ser	AAT Asn 700	AGA Arg	CAG Gln	GTG Val	GAT Asp	GAA Glu 705	CTG Leu	CAA Gln	CAA Gln	CCG Pro	CTG Leu 710	GAG Glu	CTT Leu	AAG Lys	2163
CCA Pro 715	GAG Glu	CCA Pro	GAG Glu	CTG Leu	GAG Glu	TCA Ser 720	TTA Leu	GAG Glu	CTG Leu	GAA Glu	CTA Leu 725	GGG Gly	CTG Leu	GTG Val	CCA Pro	2211

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

GAG CCA GAG CTC AGC CTG GAC TTA GAG CCA CTG CTG AAG GCA GGG CTG Glu Pro Glu Leu Ser Leu Asp Leu Glu Pro Leu Leu Lys Ala Gly Leu 730 735 740 745	2259
GAT CTG GGG CCA GAG CTA GAG TCT GTG CTG GAG TCC ACT CTG GAG CCT Asp Leu Gly Pro Glu Leu Glu Ser Val Leu Glu Ser Thr Leu Glu Pro 750 755 760	2307
GTG ATA GAG CCC ACA CTA TGC ATG GTA TCA CAA ACA GTG CCA GAG CCA Val Ile Glu Pro Thr Leu Cys Met Val Ser Gln Thr Val Pro Glu Pro 765 770 775	2355
GAC CAA GGA CCT GTA TCA CAG CCA GTG CCA GAG CCA GAT TTG CCC TGT Asp Gln Gly Pro Val Ser Gln Pro Val Pro Glu Pro Asp Leu Pro Cys 780 785 790	2403
GAT CTG AGA CAT TTG AAC ACT GAG CCA ATG GAA ATC TTC AGA AAC TGT Asp Leu Arg His Leu Asn Thr Glu Pro Met Glu Ile Phe Arg Asn Cys 795 800 805	2451
GTA AAG ATT GAA GAA ATC ATG CCG AAT GGT GAC CCA CTG TTG GCT GGC Val Lys Ile Glu Glu Ile Met Pro Asn Gly Asp Pro Leu Leu Ala Gly 810 815 820 825	2499
CAG AAC ACC GTG GAT GAG GTT TAC GTC TCC CGC CCC AGC CAC TTC TAC Gln Asn Thr Val Asp Glu Val Tyr Val Ser Arg Pro Ser His Phe Tyr 830 835 840	2547
ACT GAT GGA CCC TTG ATG CCT TCT GAC TTC TAGGAACCAC ATTTCTCTCTG Thr Asp Gly Pro Leu Met Pro Ser Asp Phe 845 850	2597
TTCTTTTCAT ATCTCTTTGC CCTTCCTACT CCTCATAGCA TGATATTGTT CTCCAAGGAT	2657
GGGAATCAGG CATGTGTCCC TTCCAAGCTG TGTAACTGT TCAAACCTCAG GCCTGTGTGA	2717
CTCCATTGGG GTGAGAGGTG AAAGCATAAC ATGGGTACAG AGGGGACAAC AATGAATCAG	2777
AACAGATGCT GAGCCATAGG TCTAAATAGG ATCCTGGAGG CTGCCCTGCTG TGCTGGGAGG	2837
TATAGGGGTC CTGGGGGCAG GCCAGGGCAG TTGACAGGTA CTTGGAGGGC TCAGGGCAGT	2897
GGCTTCTTTC CAGTATGGAA GGATTTC AAC ATTTTAATAG TTGGTTAGGC TAAACTGGTG	2957
CATACTGGCA TTGGCCTTGG TGGGGAGCAC AGACACAGGA TAGGACTCCA TTTCTTTCTT	3017
CCATTCCTTC ATGTCTAGGA TAACTTGCTT TCTTCTTTCC TTTACTCCTG GCTCAAGCCC	3077
TGAATTTCTT CTTTTCCTGC AGGGGTTGAG AGCTTTCTGC CTTAGCCTAC CATGTGAAAC	3137
TCTACCCTGA AGAAAGGGAT GGATAGGAAG TAGACCTCTT TTTCTTACCA GTCTCCTCCC	3197
CTACTCTGCC CCCTAAGCTG GCTGTACCTG TTCCTCCCC ATAAAATGAT CCTGCCAATC	3257
TAAAAAAAAA A	3268

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Gln Trp Glu Met Leu Gln Asn Leu Asp Ser Pro Phe Gln Asp
 1 5 10 15
 Gln Leu His Gln Leu Tyr Ser His Ser Leu Leu Pro Val Asp Ile Arg
 20 25 30
 Gln Tyr Leu Ala Val Trp Ile Glu Asp Gln Asn Trp Gln Glu Ala Ala
 35 40 45
 Leu Gly Ser Asp Asp Ser Lys Ala Thr Met Leu Phe Phe His Phe Leu
 50 55 60
 Asp Gln Leu Asn Tyr Glu Cys Gly Arg Cys Ser Gln Asp Pro Glu Ser
 65 70 75 80
 Leu Leu Leu Gln His Asn Leu Arg Lys Phe Cys Arg Asp Ile Gln Pro
 85 90 95
 Phe Ser Gln Asp Pro Thr Gln Leu Ala Glu Met Ile Phe Asn Leu Leu
 100 105 110
 Leu Glu Glu Lys Arg Ile Leu Ile Gln Ala Gln Arg Ala Gln Leu Glu
 115 120 125
 Gln Gly Glu Pro Val Leu Glu Thr Pro Val Glu Ser Gln Gln His Glu
 130 135 140
 Ile Glu Ser Arg Ile Leu Asp Leu Arg Ala Met Met Glu Lys Leu Val
 145 150 155 160
 Lys Ser Ile Ser Gln Leu Lys Asp Gln Gln Asp Val Phe Cys Phe Arg
 165 170 175
 Tyr Lys Ile Gln Ala Lys Gly Lys Thr Pro Ser Leu Asp Pro His Gln
 180 185 190
 Thr Lys Glu Gln Lys Ile Leu Gln Glu Thr Leu Asn Glu Leu Asp Lys
 195 200 205
 Arg Arg Lys Glu Val Leu Asp Ala Ser Lys Ala Leu Leu Gly Arg Leu
 210 215 220
 Thr Thr Leu Ile Glu Leu Leu Leu Pro Lys Leu Glu Glu Trp Lys Ala
 225 230 235 240
 Gln Gln Gln Lys Ala Cys Ile Arg Ala Pro Ile Asp His Gly Leu Glu
 245 250 255
 Gln Leu Glu Thr Trp Phe Thr Ala Gly Ala Lys Leu Leu Phe His Leu
 260 265 270
 Arg Gln Leu Leu Lys Glu Leu Lys Gly Leu Ser Cys Leu Val Ser Tyr
 275 280 285
 Gln Asp Asp Pro Leu Thr Lys Gly Val Asp Leu Arg Asn Ala Gln Val
 290 295 300
 Thr Glu Leu Leu Gln Arg Leu Leu His Arg Ala Phe Val Val Glu Thr
 305 310 315 320
 Gln Pro Cys Met Pro Gln Thr Pro His Arg Pro Leu Ile Leu Lys Thr
 325 330 335
 Gly Ser Lys Phe Thr Val Arg Thr Arg Leu Leu Val Arg Leu Gln Glu
 340 345 350
 Gly Asn Glu Ser Leu Thr Val Glu Val Ser Ile Asp Arg Asn Pro Pro
 355 360 365

Gln	Leu	Gln	Gly	Phe	Arg	Lys	Phe	Asn	Ile	Leu	Thr	Ser	Asn	Gln	Lys	370	375	380
Thr	Leu	Thr	Pro	Glu	Lys	Gly	Gln	Ser	Gln	Gly	Leu	Ile	Trp	Asp	Phe	385	390	395
Gly	Tyr	Leu	Thr	Leu	Val	Glu	Gln	Arg	Ser	Gly	Gly	Ser	Gly	Lys	Gly	405	410	415
Ser	Asn	Lys	Gly	Pro	Leu	Gly	Val	Thr	Glu	Glu	Leu	His	Ile	Ile	Ser	420	425	430
Phe	Thr	Val	Lys	Tyr	Thr	Tyr	Gln	Gly	Leu	Lys	Gln	Glu	Leu	Lys	Thr	435	440	445
Asp	Thr	Leu	Pro	Val	Val	Ile	Ile	Ser	Asn	Met	Asn	Gln	Leu	Ser	Ile	450	455	460
Ala	Trp	Ala	Ser	Val	Leu	Trp	Phe	Asn	Leu	Leu	Ser	Pro	Asn	Leu	Gln	465	470	475
Asn	Gln	Gln	Phe	Phe	Ser	Asn	Pro	Pro	Lys	Ala	Pro	Trp	Ser	Leu	Leu	485	490	495
Gly	Pro	Ala	Leu	Ser	Trp	Gln	Phe	Ser	Ser	Tyr	Val	Gly	Arg	Gly	Leu	500	505	510
Asn	Ser	Asp	Gln	Leu	Ser	Met	Leu	Arg	Asn	Lys	Leu	Phe	Gly	Gln	Asn	515	520	525
Cys	Arg	Thr	Glu	Asp	Pro	Leu	Leu	Ser	Trp	Ala	Asp	Phe	Thr	Lys	Arg	530	535	540
Glu	Ser	Pro	Pro	Gly	Lys	Leu	Pro	Phe	Trp	Thr	Trp	Leu	Asp	Lys	Ile	545	550	555
Leu	Glu	Leu	Val	His	Asp	His	Leu	Lys	Asp	Leu	Trp	Asn	Asp	Gly	Arg	565	570	575
Ile	Met	Gly	Phe	Val	Ser	Arg	Ser	Gln	Glu	Arg	Arg	Leu	Leu	Lys	Lys	580	585	590
Thr	Met	Ser	Gly	Thr	Phe	Leu	Leu	Arg	Phe	Ser	Glu	Ser	Ser	Glu	Gly	595	600	605
Gly	Ile	Thr	Cys	Ser	Trp	Val	Glu	His	Gln	Asp	Asp	Asp	Lys	Val	Leu	610	615	620
Ile	Tyr	Ser	Val	Gln	Pro	Tyr	Thr	Lys	Glu	Val	Leu	Gln	Ser	Leu	Pro	625	630	635
Leu	Thr	Glu	Ile	Ile	Arg	His	Tyr	Gln	Leu	Leu	Thr	Glu	Glu	Asn	Ile	645	650	655
Pro	Glu	Asn	Pro	Leu	Arg	Phe	Leu	Tyr	Pro	Arg	Ile	Pro	Arg	Asp	Glu	660	665	670
Ala	Phe	Gly	Cys	Tyr	Tyr	Gln	Glu	Lys	Val	Asn	Leu	Gln	Glu	Arg	Arg	675	680	685
Lys	Tyr	Leu	Lys	His	Arg	Leu	Ile	Val	Val	Ser	Asn	Arg	Gln	Val	Asp	690	695	700
Glu	Leu	Gln	Gln	Pro	Leu	Glu	Leu	Lys	Pro	Glu	Pro	Glu	Leu	Glu	Ser	705	710	715
Leu	Glu	Leu	Glu	Leu	Gly	Leu	Val	Pro	Glu	Pro	Glu	Leu	Ser	Leu	Asp	725	730	735

GAG Glu	CAC His	GCT Ala	GCC Ala	AAT Asn	GAT Asp	GTT Val	TCA Ser	TTT Phe	GCC Ala	ACC Thr	ATC Ile	CGT Arg	TTT Phe	CAT His	GAC Asp	373
45						50					55					
CTC Leu	CTG Leu	TCA Ser	CAG Gln	CTG Leu	GAT Asp	GAT Asp	CAA Gln	TAT Tyr	AGT Ser	CGC Arg	TTT Phe	TCT Ser	TTG Leu	GAG Glu	AAT Asn	421
60					65					70					75	
AAC Asn	TTC Phe	TTG Leu	CTA Leu	CAG Gln	CAT His	AAC Asn	ATA Ile	AGG Arg	AAA Lys	AGC Ser	AAG Lys	CGT Arg	AAT Asn	CTT Leu	CAG Gln	469
				80					85					90		
GAT Asp	AAT Asn	TTT Phe	CAG Gln	GAA Glu	GAC Asp	CCA Pro	ATC Ile	CAG Gln	ATG Met	TCT Ser	ATG Met	ATC Ile	ATT Ile	TAC Tyr	AGC Ser	517
			95					100					105			
TGT Cys	CTG Leu	AAG Lys	GAA Glu	GAA Glu	AGG Arg	AAA Lys	ATT Ile	CTG Leu	GAA Glu	AAC Asn	GCC Ala	CAG Gln	AGA Arg	TTT Phe	AAT Asn	565
		110					115					120				
CAG Gln	GCT Ala	CAG Gln	TCG Ser	GGG Gly	AAT Asn	ATT Ile	CAG Gln	AGC Ser	ACA Thr	GTG Val	ATG Met	TTA Leu	GAC Asp	AAA Lys	CAG Gln	613
125						130					135					
AAA Lys	GAG Glu	CTT Leu	GAC Asp	AGT Ser	AAA Lys	GTC Val	AGA Arg	AAT Asn	GTG Val	AAG Lys	GAC Asp	AAG Lys	GTT Val	ATG Met	TGT Cys	661
140					145					150					155	
ATA Ile	GAG Glu	CAT His	GAA Glu	ATC Ile	AAG Lys	AGC Ser	CTG Leu	GAA Glu	GAT Asp	TTA Leu	CAA Gln	GAT Asp	GAA Glu	TAT Tyr	GAC Asp	709
				160					165					170		
TTC Phe	AAA Lys	TGC Cys	AAA Lys	ACC Thr	TTG Leu	CAG Gln	AAC Asn	AGA Arg	GAA Glu	CAC His	GAG Glu	ACC Thr	AAT Asn	GGT Gly	GTG Val	757
			175					180					185			
GCA Ala	AAG Lys	AGT Ser	GAT Asp	CAG Gln	AAA Lys	CAA Gln	GAA Glu	CAG Gln	CTG Leu	TTA Leu	CTC Leu	AAG Lys	AAG Lys	ATG Met	TAT Tyr	805
		190					195					200				
TTA Leu	ATG Met	CTT Leu	GAC Asp	AAT Asn	AAG Lys	AGA Arg	AAG Lys	GAA Glu	GTA Val	GTT Val	CAC His	AAA Lys	ATA Ile	ATA Ile	GAG Glu	853
	205					210					215					
TTG Leu	CTG Leu	AAT Asn	GTC Val	ACT Thr	GAA Glu	CTT Leu	ACC Thr	CAG Gln	AAT Asn	GCC Ala	CTG Leu	ATT Ile	AAT Asn	GAT Asp	GAA Glu	901
220					225					230					235	
CTA Leu	GTG Val	GAG Glu	TGG Trp	AAG Lys	CGG Arg	AGA Arg	CAG Gln	CAG Gln	AGC Ser	GCC Ala	TGT Cys	ATT Ile	GGG Gly	GGG Gly	CCG Pro	949
				240					245					250		
CCC Pro	AAT Asn	GCT Ala	TGC Cys	TTG Leu	GAT Asp	CAG Gln	CTG Leu	CAG Gln	AAC Asn	TGG Trp	TTC Phe	ACT Thr	ATA Ile	GTT Val	GCG Ala	997
			255				260						265			
GAG Glu	AGT Ser	CTG Leu	CAG Gln	CAA Gln	GTT Val	CGG Arg	CAG Gln	CAG Gln	CTT Leu	AAA Lys	AAG Lys	TTG Leu	GAG Glu	GAA Glu	TTG Leu	1045
		270					275					280				
GAA Glu	CAG Gln	AAA Lys	TAC Tyr	ACC Thr	TAC Tyr	GAA Glu	CAT His	GAC Asp	CCT Pro	ATC Ile	ACA Thr	AAA Lys	AAC Asn	AAA Lys	CAA Gln	1093
	285					290					295					
GTG Val	TTA Leu	TGG Trp	GAC Asp	CGC Arg	ACC Thr	TTC Phe	AGT Ser	CTT Leu	TTC Phe	CAG Gln	CAG Gln	CTC Leu	ATT Ile	CAG Gln	AGC Ser	1141
300						305				310					315	

TCG TTT GTG GTG GAA AGA CAG CCC TGC ATG CCA ACG CAC CCT CAG AGG	1189
Ser Phe Val Val Glu Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg	
320 325 330	
CCG CTG GTC TTG AAG ACA GGG GTC CAG TTC ACT GTG AAG TTG AGA CTG	1237
Pro Leu Val Lys Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu	
335 340 345	
TTG GTG AAA TTG CAA GAG CTG AAT TAT AAT TTG AAA GTC AAA GTC TTA	1285
Leu Val Lys Leu Gln Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu	
350 355 360	
TTT GAT AAA GAT GTG AAT GAG AGA AAT ACA GTA AAA GGA TTT AGG AAG	1333
Phe Asp Lys Asp Val Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys	
365 370 375	
TTC AAC ATT TTG GGC ACG CAC ACA AAA GTG ATG AAC ATG GAG GAG TCC	1381
Phe Asn Ile Leu Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser	
380 385 390 395	
ACC AAT GGC AGT CTG GCG GCT GAA TTT CGG CAC CTG CAA TTG AAA GAA	1429
Thr Asn Gly Ser Leu Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu	
400 405 410	
CAG AAA AAT GCT GGC ACC AGA ACG AAT GAG GGT CCT CTC ATC GTT ACT	1477
Gln Lys Asn Ala Gly Thr Arg Thr Asn Glu Gly Pro Leu Ile Val Thr	
415 420 425	
GAA GAG CTT CAC TCC CTT AGT TTT GAA ACC CAA TTG TGC CAG CCT GGT	1525
Glu Glu Leu His Ser Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly	
430 435 440	
TTG GTA ATT GAC CTC GAG ACG ACC TCT CTG CCC GTT GTG GTG ATC TCC	1573
Leu Val Ile Asp Leu Glu Thr Ser Leu Pro Val Val Ile Ser	
445 450 455	
AAC GTC AGC CAG CTC CCG AGC GGT TGG GCC TCC ATC CTT TGG TAC AAC	1621
Asn Val Ser Gln Leu Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn	
460 465 470 475	
ATG CTG GTG GCG GAA CCC AGG AAT CTG TCC TTC TTC CTG ACT CCA CCA	1669
Met Leu Val Ala Glu Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro Pro	
480 485 490	
TGT GCA CGA TGG GCT CAG CTT TCA GAA GTG CTG AGT TGG CAG TTT TCT	1717
Cys Ala Arg Trp Ala Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser	
495 500 505	
TCT GTC ACC AAA AGA GGT CTC AAT GTG GAC CAG CTG AAC ATG TTG GGA	1765
Ser Val Thr Lys Arg Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly	
510 515 520	
GAG AAG CTT CTT GGT CCT AAC GCC AGC CCC GAT GGT CTC ATT CCG TGG	1813
Glu Lys Leu Leu Gly Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp	
525 530 535	
ACG AGG TTT TGT AAG GAA AAT ATA AAT GAT AAA AAT TTT CCC TTC TGG	1861
Thr Arg Phe Cys Lys Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp	
540 545 550 555	
CTT TGG ATT GAA AGC ATC CTA GAA CTC ATT AAA AAA CAC CTG CTC CCT	1909
Leu Trp Ile Glu Ser Ile Leu Glu Leu Ile Lys Lys His Leu Leu Pro	
560 565 570	
CTC TGG AAT GAT GGG TGC ATC ATG GGC TTC ATC AGC AAG GAG CGA GAG	1957
Leu Trp Asn Asp Gly Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu	
575 580 585	

CGT GCC CTG TTG AAG GAC CAG CAG CCG GGG ACC TTC CTG CTG CGG TTC Arg Ala Leu Leu Lys Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe 590 595 600	2005
AGT GAG AGC TCC CGG GAA GGG GCC ATC ACA TTC ACA TGG GTG GAG CGG Ser Glu Ser Ser Arg Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg 605 610 615	2053
TCC CAG AAC GGA GGC GAA CCT GAC TTC CAT GCG GTT GAA CCC TAC ACG Ser Gln Asn Gly Gly Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr 620 625 630 635	2101
AAG AAA GAA CTT TCT GCT GTT ACT TTC CCT GAC ATC ATT CGC AAT TAC Lys Lys Glu Leu Ser Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr 640 645 650	2149
AAA GTC ATG GCT GCT GAG AAT ATT CCT GAG AAT CCC CTG AAG TAT CTG Lys Val Met Ala Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu 655 660 665	2197
TAT CCA AAT ATT GAC AAA GAC CAT GCC TTT GGA AAG TAT TAC TCC AGG Tyr Pro Asn Ile Asp Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg 670 675 680	2245
CCA AAG GAA GCA CCA GAG CCA ATG GAA CTT GAT GGC CCT AAA GGA ACT Pro Lys Glu Ala Pro Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr 685 690 695	2293
GGA TAT ATC AAG ACT GAG TTG ATT TCT GTG TCT GAA GTT CAC CCT TCT Gly Tyr Ile Lys Thr Glu Leu Ile Ser Val Ser Glu Val His Pro Ser 700 705 710 715	2341
AGA CTT CAG ACC ACA GAC AAC CTG CTC CCC ATG TCT CCT GAG GAG TTT Arg Leu Gln Thr Thr Asp Asn Leu Leu Pro Met Ser Pro Glu Glu Phe 720 725 730	2389
GAC GAG GTG TCT CGG ATA GTG GGC TCT GTA GAA TTC GAC AGT ATG ATG Asp Glu Val Ser Arg Ile Val Gly Ser Val Glu Phe Asp Ser Met Met 735 740 745	2437
AAC ACA GTA TAGAGCATGA ATTTTTTTCA TCTTCTCTGG CGACAGTTTT Asn Thr Val 750	2485
CCTTCTCATC TGTGATTCCC TCCTGCTACT CTGTTCTTC ACATCCTGTG TTTCTAGGGA	2546
AATGAAAGAA AGGCCAGCAA ATTCGCTGCA ACCTGTTGAT AGCAAGTGAA TTTTCTCTTA	2606
ACTCAGAAAC ATCAGTTACT CTGAAGGGCA TCATGCATCT TACTGAAGGT AAAATTGAAA	2666
GGCATTCTCT GAAGAGTGGG TTTCACAAAGT GAAAAACATC CAGATACACC CAAAGTATCA	2726
GGACGAGAAT GAGGGTCCTT TGGGAAAGGA GAAGTTAAGC AACATCTAGC AAATGTTATG	2786
CATAAAGTCA GTGCCCAACT GTTATAGGTT GTTGGATAAA TCAGTGGTTA TTTAGGGAAC	2846
TGCTTGACGT AGGAACGGTA AATTCTGTG GGAGAATTCT TACATGTTTT CTTTGCTTTA	2906
AGTGTAAGTG GCAGTTTTCC ATTGGTTTAC CTGTGAAATA GTTCAAAGCC AAGTTTATAT	2966
ACAATTATAT CAGTCCTCTT TCAAAGGTAG CCATCATGGA TCTGGTAGGG GGAAAATGTG	3026
TATTTTATTA CATCTTTCAC ATTGGCTATT TAAAGACAAA GACAAATTCT GTTTCTTGAG	3086
AAGAGAACAT TTCCAAATTC ACAAGTTGTG TTTGATATCC AAAGCTGAAT ACATTCTGCT	3146
TTCATCTTGG TCACATACAA TTATTTTAC AGTTCTCCCA AGGGAGTTAG GCTATTACAA	3206
ACCACTCATT CAAAAGTTGA AATTAACCAT AGATGTAGAT AAACCTCAGAA ATTTAATTCA	3266

TGTTTCTTAA ATGGGCTACT TTGTCCTTTT TGTTATTAGG GTGGTATTTA GTCTATTAGC	3326
CACAAAATTG GGAAAGGAGT AGAAAAAGCA GTAACGACA ACTTGAATAA TACACCAGAG	3386
ATAATATGAG AATCAGATCA TTTCAAAACT CATTTCCTAT GTAACGACAT TGAGAACTGC	3446
ATATGTTTCG CTGATATATG TGTTTTTTCAC ATTTGCGAAT GGTTCATTC TCTCTCCTGT	3506
ACTTTTTCCA GACACTTTTT TGAGTGGATG ATGTTTCGTG AAGTATACTG TATTTTTACC	3566
TTTTTCCTTC CTTATCACTG ACACAAAAG TAGATTAAGA GATGGGTTTG ACAAGGTTCT	3626
TCCCTTTTAC AACTGCTGT CTATGTGGCT GTATCTTGTT TTTCCACTAC TGCTACCACA	3686
ACTATAATTAT CATGCAAATG CTGTATTCTT CTTGGTGGA GATAAAGATT TCTTGAGTTT	3746
TGTTTTAAAA TTAAAGCTAA AGTATCTGTA TTGCATTAAA TATAATATCG ACACAGTGCT	3806
TTCCGTGGCA CTGCATACAA TCTGAGGCCT CCTCTCTCAG TTTTATATA GATGGCGAGA	3866
ACCTAAGTTT CAGTTGATTT TACAATTGAA ATGACTAAAA AACAAAGAAG ACAACATTAA	3926
AAACAATATT GTTTCTA	3943

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 750 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Ser	Gln	Trp	Tyr	Glu	Leu	Gln	Gln	Leu	Asp	Ser	Lys	Phe	Leu	Glu
1				5					10					15	
Gln	Val	His	Gln	Leu	Tyr	Asp	Asp	Ser	Phe	Pro	Met	Glu	Ile	Arg	Gln
			20					25					30		
Tyr	Leu	Ala	Gln	Trp	Leu	Glu	Lys	Gln	Asp	Trp	Glu	His	Ala	Ala	Asn
		35					40					45			
Asp	Val	Ser	Phe	Ala	Thr	Ile	Arg	Phe	His	Asp	Leu	Leu	Ser	Gln	Leu
	50					55					60				
Asp	Asp	Gln	Tyr	Ser	Arg	Phe	Ser	Leu	Glu	Asn	Asn	Phe	Leu	Leu	Gln
65				70						75				80	
His	Asn	Ile	Arg	Lys	Ser	Lys	Arg	Asn	Leu	Gln	Asp	Asn	Phe	Gln	Glu
			85					90						95	
Asp	Pro	Ile	Gln	Met	Ser	Met	Ile	Ile	Tyr	Ser	Cys	Leu	Lys	Glu	Glu
		100					105						110		
Arg	Lys	Ile	Leu	Glu	Asn	Ala	Gln	Arg	Phe	Asn	Gln	Ala	Gln	Ser	Gly
		115				120						125			
Asn	Ile	Gln	Ser	Thr	Val	Met	Leu	Asp	Lys	Gln	Lys	Glu	Leu	Asp	Ser
130						135					140				
Lys	Val	Arg	Asn	Val	Lys	Asp	Lys	Val	Met	Cys	Ile	Glu	His	Glu	Ile
145				150						155				160	
Lys	Ser	Leu	Glu	Asp	Leu	Gln	Asp	Glu	Tyr	Asp	Phe	Lys	Cys	Lys	Thr
			165					170						175	

Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val Ala Lys Ser Asp Gln
 180 185 190
 Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr Leu Met Leu Asp Asn
 195 200 205
 Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu Leu Asn Val Thr
 210 215 220
 Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu Trp Lys
 225 230 235 240
 Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
 245 250 255
 Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln Gln
 260 265 270
 Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr Thr
 275 280 285
 Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp Arg
 290 295 300
 Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val Glu
 305 310 315 320
 Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu Lys
 325 330 335
 Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu Leu Val Lys Leu Gln
 340 345 350
 Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe Asp Lys Asp Val
 355 360 365
 Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile Leu Gly
 370 375 380
 Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser Leu
 385 390 395 400
 Ala Ala Glu Phe Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala Gly
 405 410 415
 Thr Arg Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His Ser
 420 425 430
 Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp Leu
 435 440 445
 Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser Asn Val Ser Gln Leu
 450 455 460
 Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn Met Leu Val Ala Glu
 465 470 475 480
 Pro Arg Asn Leu Ser Phe Phe Leu Thr Pro Pro Cys Ala Arg Trp Ala
 485 490 495
 Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser Val Thr Lys Arg
 500 505 510
 Gly Leu Asn Val Asp Gln Leu Asn Met Leu Gly Glu Lys Leu Leu Gly
 515 520 525
 Pro Asn Ala Ser Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe Cys Lys
 530 535 540

Glu Asn Ile Asn Asp Lys Asn Phe Pro Phe Trp Leu Trp Ile Glu Ser
 545 550 555
 Ile Leu Glu Leu Ile Lys Lys His Leu Leu Pro Leu Trp Asn Asp Gly
 565 570 575
 Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu Leu Lys
 580 585 590
 Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser Ser Arg
 595 600 605
 Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg Ser Gln Asn Gly Gly
 610 615 620
 Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr Lys Lys Glu Leu Ser
 625 630 635 640
 Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys Val Met Ala Ala
 645 650 655
 Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn Ile Asp
 660 665 670
 Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu Ala Pro
 675 680 685
 Glu Pro Met Glu Leu Asp Gly Pro Lys Gly Thr Gly Tyr Ile Lys Thr
 690 695 700
 Glu Leu Ile Ser Val Ser Glu Val His Pro Ser Arg Leu Gln Thr Thr
 705 710 715 720
 Asp Asn Leu Leu Pro Met Ser Pro Glu Glu Phe Asp Glu Val Ser Arg
 725 730 735
 Ile Val Gly Ser Val Glu Phe Asp Ser Met Met Asn Thr Val
 740 745 750

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2607 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 197..2335

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTAAACCTC TCGCCGAGCC CCTCCGCAGA CTCTGCGCCG GAAAGTTTCA TTTGCTGTAT	60
GCCATCCTCG AGAGCTGTCT AGGTTAACGT TCGCACTCTG TGTATATAAC CTCGACAGTC	120
TTGGCACCTA ACGTGCTGTG CGTAGCTGCT CCTTTGGTTG AATCCCCAGG CCCTTGTTGG	180

GGCACAAAGGT GGCAGG ATG TCT CAG TGG TAC GAA CTT CAG CAG CTT GAC	229
Met Ser Gln Trp Tyr Glu Leu Gln Gln Leu Asp	
1 5 10	
TCA AAA TTC CTG GAG CAG GTT CAC CAG CTT TAT GAT GAC AGT TTT CCC	277
Ser Lys Phe Leu Glu Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro	
15 20 25	
ATG GAA ATC AGA CAG TAC CTG GCA CAG TGG TTA GAA AAG CAA GAC TGG	325
Met Glu Ile Arg Gln Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp	
30 35 40	
GAG CAC GCT GCC AAT GAT GTT TCA TTT GCC ACC ATC CGT TTT CAT GAC	373
Glu His Ala Ala Asn Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp	
45 50 55	
CTC CTG TCA CAG CTG GAT GAT CAA TAT AGT CGC TTT TCT TTG GAG AAT	421
Leu Leu Ser Gln Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn	
60 65 70 75	
AAC TTC TTG CTA CAG CAT AAC ATA AGG AAA AGC AAG CGT AAT CTT CAG	469
Asn Phe Leu Leu Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln	
80 85 90	
GAT AAT TTT CAG GAA GAG CCA ATC CAG ATG TCT ATG ATC ATT TAC AGC	517
Asp Asn Phe Gln Glu Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser	
95 100 105	
TGT CTG AAG GAA GAA AGG AAA ATT CTG GAA AAC GCC CAG AGA TTT AAT	565
Cys Leu Lys Glu Glu Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn	
110 115 120	
CAG GCT CAG TCG GGG AAT ATT CAG AGC ACA GTG ATG TTA GAC AAA CAG	613
Gln Ala Gln Ser Gly Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln	
125 130 135	
AAA GAG CTT GAC AGT AAA GTC AGA AAT GTG AAG GAC AAG GTT ATG TGT	661
Lys Glu Leu Asp Ser Lys Val Arg Asn Val Lys Asp Lys Val Met Cys	
140 145 150 155	
ATA GAG CAT GAA ATC AAG AGC CTG GAA GAT TTA CAA GAT GAA TAT GAC	709
Ile Glu His Glu Ile Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp	
160 165 170	
TTC AAA TGC AAA ACC TTG CAG AAC AGA GAA CAC GAG ACC AAT GGT GTG	757
Phe Lys Cys Lys Thr Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val	
175 180 185	
GCA AAG AGT GAT CAG AAA CAA GAA CAG CTG TTA CTC AAG AAG ATG TAT	805
Ala Lys Ser Asp Gln Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr	
190 195 200	
TTA ATG CTT GAC AAT AAG AGA AAG GAA GTA GTT CAC AAA ATA ATA GAG	853
Leu Met Leu Asp Asn Lys Arg Lys Glu Val Val His Lys Ile Ile Glu	
205 210 215	
TTG CTG AAT GTC ACT GAA CTT ACC CAG AAT GCC CTG ATT AAT GAT GAA	901
Leu Leu Asn Val Thr Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu	
220 225 230 235	
CTA GTG GAG TGG AAG CGG AGA CAG CAG AGC GCC TGT ATT GGG GGG CCG	949
Leu Val Glu Trp Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro	
240 245 250	
CCC AAT GCT TGC TTG GAT CAG CTG CAG AAC TGG TTC ACT ATA GTT GCG	997
Pro Asn Ala Cys Leu Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala	
255 260 265	
GAG AGT CTG CAG CAA GTT CGG CAG CAG CTT AAA AAG TTG GAG GAA TTG	1045

Glu	Ser	Leu	Gln	Gln	Val	Arg	Gln	Gln	Leu	Lys	Lys	Leu	Glu	Glu	Leu		
		270					275					280					
GAA	CAG	AAA	TAC	ACC	TAC	GAA	CAT	GAC	CCT	ATC	ACA	AAA	AAC	AAA	CAA	1093	
Glu	Gln	Lys	Tyr	Thr	Tyr	Glu	His	Asp	Pro	Ile	Thr	Lys	Asn	Lys	Gln		
		285				290				295							
GTG	TTA	TGG	GAC	CGC	ACC	TTC	AGT	CTT	TTC	CAG	CAG	CTC	ATT	CAG	AGC	1141	
Val	Leu	Trp	Asp	Arg	Thr	Phe	Ser	Leu	Phe	Gln	Gln	Leu	Ile	Gln	Ser		
		300			305					310					315		
TCG	TTT	GTG	GTG	GAA	AGA	CAG	CCC	TGC	ATG	CCA	ACG	CAC	CCT	CAG	AGG	1189	
Ser	Phe	Val	Val	Glu	Arg	Gln	Pro	Cys	Met	Pro	Thr	His	Pro	Gln	Arg		
				320					325					330			
CCG	CTG	GTC	TTG	AAG	ACA	GGG	GTC	CAG	TTC	ACT	GTG	AAG	TTG	AGA	CTG	1237	
Pro	Leu	Val	Leu	Lys	Thr	Gly	Val	Gln	Phe	Thr	Val	Lys	Leu	Arg	Leu		
			335					340					345				
TTG	GTG	AAA	TTG	CAA	GAG	CTG	AAT	TAT	AAT	TTG	AAA	GTC	AAA	GTC	TTA	1285	
Leu	Val	Lys	Leu	Gln	Glu	Leu	Asn	Tyr	Asn	Leu	Lys	Val	Lys	Val	Leu		
		350				355						360					
TTT	GAT	AAA	GAT	GTG	AAT	GAG	AGA	AAT	ACA	GTA	AAA	GGA	TTT	AGG	AAG	1333	
Phe	Asp	Lys	Asp	Val	Asn	Glu	Arg	Asn	Thr	Val	Lys	Gly	Phe	Arg	Lys		
		365				370					375						
TTC	AAC	ATT	TTG	GGC	ACG	CAC	ACA	AAA	GTG	ATG	AAC	ATG	GAG	GAG	TCC	1381	
Phe	Asn	Ile	Leu	Gly	Thr	His	Thr	Lys	Val	Met	Asn	Met	Glu	Glu	Ser		
					385					390					395		
ACC	AAT	GGC	AGT	CTG	GCG	GCT	GAA	TTT	CGG	CAC	CTG	CAA	TTG	AAA	GAA	1429	
Thr	Asn	Gly	Ser	Leu	Ala	Ala	Glu	Phe	Arg	His	Leu	Gln	Leu	Lys	Glu		
				400					405					410			
CAG	AAA	AAT	GCT	GGC	ACC	AGA	ACG	AAT	GAG	GGT	CCT	CTC	ATC	GTT	ACT	1477	
Gln	Lys	Asn	Ala	Gly	Thr	Arg	Thr	Asn	Glu	Gly	Pro	Leu	Ile	Val	Thr		
			415					420					425				
GAA	GAG	CTT	CAC	TCC	CTT	AGT	TTT	GAA	ACC	CAA	TTG	TGC	CAG	CCT	GGT	1525	
Glu	Glu	Leu	His	Ser	Leu	Ser	Phe	Glu	Thr	Gln	Leu	Cys	Gln	Pro	Gly		
		430					435					440					
TTG	GTA	ATT	GAC	CTC	GAG	ACG	ACC	TCT	CTG	CCC	GTT	GTG	GTG	ATC	TCC	1573	
Leu	Val	Ile	Asp	Leu	Glu	Thr	Thr	Ser	Leu	Pro	Val	Val	Val	Ile	Ser		
		445				450					455						
AAC	GTC	AGC	CAG	CTC	CCG	AGC	GGT	TGG	GCC	TCC	ATC	CTT	TGG	TAC	AAC	1621	
Asn	Val	Ser	Gln	Leu	Pro	Ser	Gly	Trp	Ala	Ser	Ile	Leu	Trp	Tyr	Asn		
					465					470					475		
ATG	CTG	GTG	GCG	GAA	CCC	AGG	AAT	CTG	TCC	TTC	TTC	CTG	ACT	CCA	CCA	1669	
Met	Leu	Val	Ala	Glu	Pro	Arg	Asn	Leu	Ser	Phe	Phe	Leu	Thr	Pro	Pro		
				480				485						490			
TGT	GCA	CGA	TGG	GCT	CAG	CTT	TCA	GAA	GTG	CTG	AGT	TGG	CAG	TTT	TCT	1717	
Cys	Ala	Arg	Trp	Ala	Gln	Leu	Ser	Glu	Val	Leu	Ser	Trp	Gln	Phe	Ser		
			495					500					505				
TCT	GTC	ACC	AAA	AGA	GGT	CTC	AAT	GTG	GAC	CAG	CTG	AAC	ATG	TTG	GGA	1765	
Ser	Val	Thr	Lys	Arg	Gly	Leu	Asn	Val	Asp	Gln	Leu	Asn	Met	Leu	Gly		
			510			515						520					
GAG	AAG	CTT	CTT	GGT	CCT	AAC	GCC	AGC	CCC	GAT	GGT	CTC	ATT	CCG	TGG	1813	
Glu	Lys	Leu	Leu	Gly	Pro	Asn	Ala	Ser	Pro	Asp	Gly	Leu	Ile	Pro	Trp		
		525				530					535						

ACG Thr 540	AGG Arg	TTT Phe	TGT Cys	AAG Lys	GAA Glu 545	AAT Asn	ATA Ile	AAT Asn	GAT Asp	AAA Lys 550	AAT Asn	TTT Phe	CCC Pro	TTC Phe	TGG Trp 555	1861
CTT Leu	TGG Trp	ATT Ile	GAA Glu	AGC Ser 560	ATC Ile	CTA Leu	GAA Glu	CTC Leu	ATT Ile 565	AAA Lys	AAA Lys	CAC His	CTG Leu	CTC Leu 570	CCT Pro	1909
CTC Leu	TGG Trp	AAT Asn	GAT Asp 575	GGG Gly	TGC Cys	ATC Ile	ATG Met	GGC Gly 580	TTC Phe	ATC Ile	AGC Ser	AAG Lys	GAG Glu 585	CGA Arg	GAG Glu	1957
CGT Arg	GCC Ala	CTG Leu 590	TTG Leu	AAG Lys	GAC Asp	CAG Gln	CAG Gln 595	CCG Pro	GGG Gly	ACC Thr	TTC Phe	CTG Leu 600	CTG Leu	CGG Arg	TTC Phe	2005
AGT Ser 605	GAG Glu	AGC Ser	TCC Ser	CGG Arg	GAA Glu	GGG Gly 610	GCC Ala	ATC Ile	ACA Thr	TTC Phe 615	ACA Thr	TGG Trp	GTG Val	GAG Glu	CGG Arg	2053
TCC Ser 620	CAG Gln	AAC Asn	GGA Gly	GGC Gly	GAA Glu 625	CCT Pro	GAC Asp	TTC Phe	CAT His	GCG Ala 630	GTT Val	GAA Glu	CCC Pro	TAC Tyr 635	ACG Thr	2101
AAG Lys	AAA Lys	GAA Glu	CTT Leu	TCT Ser 640	GCT Ala	GTT Val	ACT Thr	TTC Phe	CCT Pro 645	GAC Asp	ATC Ile	ATT Ile	CGC Arg	AAT Asn 650	TAC Tyr	2149
AAA Lys	GTC Val	ATG Met	GCT Ala 655	GCT Ala	GAG Glu	AAT Asn	ATT Ile 660	CCT Pro	GAG Glu	AAT Asn	CCC Pro	CTG Leu 665	AAG Lys	TAT Tyr	CTG Leu	2197
TAT Tyr	CCA Pro	AAT Asn 670	ATT Ile	GAC Asp	AAA Lys	GAC Asp	CAT His 675	GCC Ala	TTT Phe	GGA Gly	AAG Lys	TAT Tyr 680	TAC Tyr	TCC Ser	AGG Arg	2245
CCA Pro 685	AAG Lys	GAA Glu	GCA Ala	CCA Pro	GAG Glu	CCA Pro 690	ATG Met	GAA Glu	CTT Leu	GAT Asp	GGC Gly 695	CCT Pro	AAA Lys	GGA Gly	ACT Thr	2293
GGA Gly 700	TAT Tyr	ATC Ile	AAG Lys	ACT Thr	GAG Glu 705	TTG Leu	ATT Ile	TCT Ser	GTG Val 710	TCT Ser	GAA Glu	GTG Val	TAAGTGAACA			2342
CAGAAGAGTG	ACATGTTTAC	AAACCTCAAG	CCAGCCTTGC	TCCTGGCTGG	GGCCTGTTGA											2402
AGATGCTTGT	ATTTTACTTT	TCCATTGTAA	TTGCTATCGC	CATCACAGCT	GAACCTGTTG											2462
AGATCCCCGT	GTTACTGCCT	ATCAGCATTT	TACTACTTTA	AAAAAAAAAA	AAAAAGCCAA											2522
AAACCAAATT	TGTATTTAAG	GTATATAAAT	TTTCCCAAAA	CTGATACCCT	TTGAAAAAGT											2582
ATAAATAAAA	TGAGCAAAAG	TTGAA														2607

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 712 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Ser	Gln	Trp	Tyr	Glu	Leu	Gln	Gln	Leu	Asp	Ser	Lys	Phe	Leu	Glu
1				5					10						15

Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile Arg Gln
 20 25 30
 Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala Ala Asn
 35 40 45
 Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser Gln Leu
 50 55 60
 Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu Leu Gln
 65 70 75 80
 His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe Gln Glu
 85 90 95
 Asp Pro Ile Gln Met Ser Met Ile Ile Tyr Ser Cys Leu Lys Glu Glu
 100 105 110
 Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Ser Gly
 115 120 125
 Asn Ile Gln Ser Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp Ser
 130 135 140
 Lys Val Arg Asn Val Lys Asp Lys Val Met Cys Ile Glu His Glu Ile
 145 150 155 160
 Lys Ser Leu Glu Asp Leu Gln Asp Glu Tyr Asp Phe Lys Cys Lys Thr
 165 170 175
 Leu Gln Asn Arg Glu His Glu Thr Asn Gly Val Ala Lys Ser Asp Gln
 180 185 190
 Lys Gln Glu Gln Leu Leu Leu Lys Lys Met Tyr Leu Met Leu Asp Asn
 195 200 205
 Lys Arg Lys Glu Val Val His Lys Ile Ile Glu Leu Leu Asn Val Thr
 210 215 220
 Glu Leu Thr Gln Asn Ala Leu Ile Asn Asp Glu Leu Val Glu Trp Lys
 225 230 235 240
 Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
 245 250 255
 Asp Gln Leu Gln Asn Trp Phe Thr Ile Val Ala Glu Ser Leu Gln Gln
 260 265 270
 Val Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Tyr Thr
 275 280 285
 Tyr Glu His Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Trp Asp Arg
 290 295 300
 Thr Phe Ser Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val Glu
 305 310 315 320
 Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu Lys
 325 330 335
 Thr Gly Val Gln Phe Thr Val Lys Leu Arg Leu Leu Val Lys Leu Gln
 340 345 350
 Glu Leu Asn Tyr Asn Leu Lys Val Lys Val Leu Phe Asp Lys Asp Val
 355 360 365
 Asn Glu Arg Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile Leu Gly
 370 375 380

Thr	His	Thr	Lys	Val	Met	Asn	Met	Glu	Glu	Ser	Thr	Asn	Gly	Ser	Leu	385	390	395	400
Ala	Ala	Glu	Phe	Arg	His	Leu	Gln	Leu	Lys	Glu	Gln	Lys	Asn	Ala	Gly	405	410	415	
Thr	Arg	Thr	Asn	Glu	Gly	Pro	Leu	Ile	Val	Thr	Glu	Glu	Leu	His	Ser	420	425	430	
Leu	Ser	Phe	Glu	Thr	Gln	Leu	Cys	Gln	Pro	Gly	Leu	Val	Ile	Asp	Leu	435	440	445	
Glu	Thr	Thr	Ser	Leu	Pro	Val	Val	Val	Ile	Ser	Asn	Val	Ser	Gln	Leu	450	455	460	
Pro	Ser	Gly	Trp	Ala	Ser	Ile	Leu	Trp	Tyr	Asn	Met	Leu	Val	Ala	Glu	465	470	475	480
Pro	Arg	Asn	Leu	Ser	Phe	Phe	Leu	Thr	Pro	Pro	Cys	Ala	Arg	Trp	Ala	485	490	495	
Gln	Leu	Ser	Glu	Val	Leu	Ser	Trp	Gln	Phe	Ser	Ser	Val	Thr	Lys	Arg	500	505	510	
Gly	Leu	Asn	Val	Asp	Gln	Leu	Asn	Met	Leu	Gly	Glu	Lys	Leu	Leu	Gly	515	520	525	
Pro	Asn	Ala	Ser	Pro	Asp	Gly	Leu	Ile	Pro	Trp	Thr	Arg	Phe	Cys	Lys	530	535	540	
Glu	Asn	Ile	Asn	Asp	Lys	Asn	Phe	Pro	Phe	Trp	Leu	Trp	Ile	Glu	Ser	545	550	555	560
Ile	Leu	Glu	Leu	Ile	Lys	Lys	His	Leu	Leu	Pro	Leu	Trp	Asn	Asp	Gly	565	570	575	
Cys	Ile	Met	Gly	Phe	Ile	Ser	Lys	Glu	Arg	Glu	Arg	Ala	Leu	Leu	Lys	580	585	590	
Asp	Gln	Gln	Pro	Gly	Thr	Phe	Leu	Leu	Arg	Phe	Ser	Glu	Ser	Ser	Arg	595	600	605	
Glu	Gly	Ala	Ile	Thr	Phe	Thr	Trp	Val	Glu	Arg	Ser	Gln	Asn	Gly	Gly	610	615	620	
Glu	Pro	Asp	Phe	His	Ala	Val	Glu	Pro	Tyr	Thr	Lys	Lys	Glu	Leu	Ser	625	630	635	640
Ala	Val	Thr	Phe	Pro	Asp	Ile	Ile	Arg	Asn	Tyr	Lys	Val	Met	Ala	Ala	645	650	655	
Glu	Asn	Ile	Pro	Glu	Asn	Pro	Leu	Lys	Tyr	Leu	Tyr	Pro	Asn	Ile	Asp	660	665	670	
Lys	Asp	His	Ala	Phe	Gly	Lys	Tyr	Tyr	Ser	Arg	Pro	Lys	Glu	Ala	Pro	675	680	685	
Glu	Pro	Met	Glu	Leu	Asp	Gly	Pro	Lys	Gly	Thr	Gly	Tyr	Ile	Lys	Thr	690	695	700	
Glu	Leu	Ile	Ser	Val	Ser	Glu	Val									705	710		

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2277 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: both
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mouse

(vii) IMMEDIATE SOURCE:

(B) CLONE: Murine Stat91

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 5..2251

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGG ATG TCA CAG TGG TTC GAG CTT CAG CAG CTG GAC TCC AAG TTC CTG	49
Met Ser Gln Trp Phe Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu	
1 5 10 15	
GAG CAG GTC CAC CAG CTG TAC GAT GAC AGT TTC CCC ATG GAA ATC AGA	97
Glu Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile Arg	
20 25 30	
CAG TAC CTG GCC CAG TGG CTG GAA AAG CAA GAC TGG GAG CAC GCT GCC	145
Gln Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala Ala	
35 40 45	
TAT GAT GTC TCG TTT GCG ACC ATC CGC TTC CAT GAC CTC CTC TCA CAG	193
Tyr Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser Gln	
50 55 60	
CTG GAC GAC CAG TAC AGC CGC TTT TCT CTG GAG AAT AAT TTC TTG TTG	241
Leu Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu Leu	
65 70 75	
CAG CAC AAC ATA CGG AAA AGC AAG CGT AAT CTC CAG GAT AAC TTC CAA	289
Gln His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe Gln	
80 85 90 95	
GAA GAT CCC GTA CAG ATG TCC ATG ATC ATC TAC AAC TGT CTG AAG GAA	337
Glu Asp Pro Val Gln Met Ser Met Ile Ile Tyr Asn Cys Leu Lys Glu	
100 105 110	
GAA AGG AAG ATT TTG GAA AAT GCC CAA AGA TTT AAT CAG GCC CAG GAG	385
Glu Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Glu	
115 120 125	
GGA AAT ATT CAG AAC ACT GTG ATG TTA GAT AAA CAG AAG GAG CTG GAC	433
Gly Asn Ile Gln Asn Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp	
130 135 140	
AGT AAA GTC AGA AAT GTG AAG GAT CAA GTC ATG TGC ATA GAG CAG GAA	481
Ser Lys Val Arg Asn Val Lys Asp Gln Val Met Cys Ile Glu Gln Glu	
145 150 155	
ATC AAG ACC CTA GAA GAA TTA CAA GAT GAA TAT GAC TTT AAA TGC AAA	529
Ile Lys Thr Leu Glu Glu Leu Gln Asp Glu Tyr Asp Phe Lys Cys Lys	
160 165 170 175	
ACC TCT CAG AAC AGA GAA GGT GAA GCC AAT GGT GTG GCG AAG AGC GAC	577
Thr Ser Gln Asn Arg Glu Gly Glu Ala Asn Gly Val Ala Lys Ser Asp	
180 185 190	

CAA AAA CAG GAA CAG CTG CTG CTC CAC AAG ATG TTT TTA ATG CTT GAC Gln Lys Gln Glu Gln Leu Leu Leu His Lys Met Phe Leu Met Leu Asp 195 200 205	625
AAT AAG AGA AAG GAG ATA ATT CAC AAA ATC AGA GAG TTG CTG AAT TCC Asn Lys Arg Lys Glu Ile Ile His Lys Ile Arg Glu Leu Leu Asn Ser 210 215 220	673
ATC GAG CTC ACT CAG AAC ACT CTG ATT AAT GAC GAG CTC GTG GAG TGG Ile Glu Leu Thr Gln Asn Thr Leu Ile Asn Asp Glu Leu Val Glu Trp 225 230 235	721
AAG CGA AGG CAG CAG AGC GCC TGC ATC GGG GGA CCG CCC AAC GCC TGC Lys Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys 240 245 250 255	769
CTG GAT CAG CTG CAA ACG TGG TTC ACC ATT GTT GCA GAG ACC CTG CAG Leu Asp Gln Leu Gln Thr Trp Phe Thr Ile Val Ala Glu Thr Leu Gln 260 265 270	817
CAG ATC CGT CAG CAG CTT AAA AAG CTG GAG GAG TTG GAA CAG AAA TTC Gln Ile Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Phe 275 280 285	865
ACC TAT GAG CCC GAC CQT ATT ACA AAA AAC AAG CAG GTG TTG TCA GAT Thr Tyr Glu Pro Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Ser Asp 290 295 300	913
CGA ACC TTC CTC CTC TTC CAG CAG CTC ATT CAG AGC TCC TTC GTG GTA Arg Thr Phe Leu Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val 305 310 315	961
GAA CGA CAG CCG TGC ATG CCC ACT CAC CCG CAG AGG CCC CTG GTC TTG Glu Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu 320 325 330 335	1009
AAG ACT GGG GTA CAG TTC ACT GTC AAG TCG AGA CTG TTG GTG AAA TTG Lys Thr Gly Val Gln Phe Thr Val Lys Ser Arg Leu Leu Val Lys Leu 340 345 350	1057
CAA GAG TCG AAT CTA TTA ACG AAA GTG AAA TGT CAC TTT GAC AAA GAT Gln Glu Ser Asn Leu Leu Thr Lys Val Lys Cys His Phe Asp Lys Asp 355 360 365	1105
GTG AAC GAG AAA AAC ACA GTT AAA GGA TTT CGG AAG TTC AAC ATC TTG Val Asn Glu Lys Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile Leu 370 375 380	1153
GGT ACG CAC ACA AAA GTG ATG AAC ATG GAA GAA TCC ACC AAC GGA AGT Gly Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser 385 390 395	1201
CTG GCA GCT GAG CTC CGA CAC CTG CAA CTG AAG GAA CAG AAA AAC GCT Leu Ala Ala Glu Leu Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala 400 405 410 415	1249
GGG AAC AGA ACT AAT GAG GGG CCT CTC ATT GTC ACC GAA GAA CTT CAC Gly Asn Arg Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His 420 425 430	1297
TCT CTT AGC TTT GAA ACC CAG TTG TGC CAG CCA GGC TTG GTG ATT GAC Ser Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp 435 440 445	1345
CTG GAG ACC ACC TCT CTT CCT GTC GTG GTG ATC TCC AAC GTC AGC CAG Leu Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser Asn Val Ser Gln 450 455 460	1393

CTC Leu	CCC Pro	AGT Ser	GGC Gly	TGG Trp	GCG Ala	TCT Ser	ATC Ile	CTG Leu	TGG Trp	TAC Tyr	AAC Asn	ATG Met	CTG Leu	GTG Val	ACA Thr	1441
465						470					475					
GAG Glu	CCC Pro	AGG Arg	AAT Asn	CTC Leu	TCC Ser	TTC Phe	TTC Phe	CTG Leu	AAC Asn	CCC Pro	CCG Pro	TGC Cys	GCG Ala	TGG Trp	TGG Trp	1489
480					485					490					495	
TCC Ser	CAG Gln	CTC Leu	TCA Ser	GAG Glu	GTG Val	TTG Leu	AGT Ser	TGG Trp	CAG Gln	TTT Phe	TCA Ser	TCA Ser	GTC Val	ACC Thr	AAG Lys	1537
				500					505					510		
AGA Arg	GGT Gly	CTG Leu	AAC Asn	GCA Ala	GAC Asp	CAG Gln	CTG Leu	AGC Ser	ATG Met	CTG Leu	GGA Gly	GAG Glu	AAG Lys	CTG Leu	CTG Leu	1585
			515					520					525			
GGC Gly	CCT Pro	AAT Asn	GCT Ala	GGC Gly	CCT Pro	GAT Asp	GGT Gly	CTT Leu	ATT Ile	CCA Pro	TGG Trp	ACA Thr	AGG Arg	TTT Phe	TGT Cys	1633
		530					535					540				
AAG Lys	GAA Glu	AAT Asn	ATT Ile	AAT Asn	GAT Asp	AAA Lys	AAT Asn	TTC Phe	TCC Ser	TTC Phe	TGG Trp	CCT Pro	TGG Trp	ATT Ile	GAC Asp	1681
	545					550					555					
ACC Thr	ATC Ile	CTA Leu	GAG Glu	CTC Leu	ATT Ile	AAG Lys	AAC Asn	GAC Asp	CTG Leu	CTG Leu	TGC Cys	CTC Leu	TGG Trp	AAT Asn	GAT Asp	1729
560					565					570					575	
GGG Gly	TGC Cys	ATT Ile	ATG Met	GGC Gly	TTC Phe	ATC Ile	AGC Ser	AAG Lys	GAG Glu	CGA Arg	GAA Glu	CGC Arg	GCT Ala	CTG Leu	CTC Leu	1777
				580					585					590		
AAG Lys	GAC Asp	CAG Gln	CAG Gln	CCA Pro	GGG Gly	ACG Thr	TTC Phe	CTG Leu	CTT Leu	AGA Arg	TTC Phe	AGT Ser	GAG Glu	AGC Ser	TCC Ser	1825
			595					600					605			
CGG Arg	GAA Glu	GGG Gly	GCC Ala	ATC Ile	ACA Thr	TTC Phe	ACA Thr	TGG Trp	GTG Val	GAA Glu	CGG Arg	TCC Ser	CAG Gln	AAC Asn	GGA Gly	1873
		610				615						620				
GGT Gly	GAA Glu	CCT Pro	GAC Asp	TTC Phe	CAT His	GCC Ala	GTG Val	GAG Glu	CCC Pro	TAC Tyr	ACG Thr	AAA Lys	AAA Lys	GAA Glu	CTT Leu	1921
	625					630					635					
TCA Ser	GCT Ala	GTT Val	ACT Thr	TTC Phe	CCA Pro	GAT Asp	ATT Ile	ATT Ile	CGC Arg	AAC Asn	TAC Tyr	AAA Lys	GTC Val	ATG Met	GCT Ala	1969
	640				645					650					655	
GCC Ala	GAG Glu	AAC Asn	ATA Ile	CCA Pro	GAG Glu	AAT Asn	CCC Pro	CTG Leu	AAG Lys	TAT Tyr	CTG Leu	TAC Tyr	CCC Pro	AAT Asn	ATT Ile	2017
				660					665					670		
GAC Asp	AAA Lys	GAC Asp	CAC His	GCC Ala	TTT Phe	GGG Gly	AAG Lys	TAT Tyr	TCC Ser	AGA Arg	CCA Pro	AAG Lys	GAA Glu	GCA Ala		2065
			675				680					685				
CCA Pro	GAA Glu	CCG Met	ATG Glu	GAG Leu	CTT Asp	GAC Asp	GAC Asp	CCT Pro	AAG Lys	CGA Arg	ACT Thr	GGA Gly	TAC Tyr	ATC Ile	AAG Lys	2113
		690				695						700				
ACT Thr	GAG Glu	TTG Leu	ATT Ile	TCT Ser	GTG Val	TCT Ser	GAA Glu	GTC Val	CAC His	CCT Pro	TCT Ser	AGA Arg	CTT Leu	CAG Gln	ACC Thr	2161
	705					710					715					
ACA Thr	GAC Asp	AAC Asn	CTG Leu	CTT Leu	CCC Pro	ATG Met	TCT Ser	CCA Pro	GAG Glu	GAG Glu	TTT Phe	GAT Asp	GAG Glu	ATG Met	TCC Ser	2209
	720				725					730					735	

CGG ATA GTG GGC CCC GAA TTT GAC AGT ATG ATG AGC ACA GTA
Arg Ile Val Gly Pro Glu Phe Asp Ser Met Met Ser Thr Val
740 745

2251

TAAACACGAA TTTCTCTCTG GCGACA

2277

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 749 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Ser Gln Trp Phe Glu Leu Gln Gln Leu Asp Ser Lys Phe Leu Glu
1 5 10 15
Gln Val His Gln Leu Tyr Asp Asp Ser Phe Pro Met Glu Ile Arg Gln
20 25 30
Tyr Leu Ala Gln Trp Leu Glu Lys Gln Asp Trp Glu His Ala Ala Tyr
35 40 45
Asp Val Ser Phe Ala Thr Ile Arg Phe His Asp Leu Leu Ser Gln Leu
50 55 60
Asp Asp Gln Tyr Ser Arg Phe Ser Leu Glu Asn Asn Phe Leu Leu Gln
65 70 75 80
His Asn Ile Arg Lys Ser Lys Arg Asn Leu Gln Asp Asn Phe Gln Glu
85 90 95
Asp Pro Val Gln Met Ser Met Ile Ile Tyr Asn Cys Leu Lys Glu Glu
100 105 110
Arg Lys Ile Leu Glu Asn Ala Gln Arg Phe Asn Gln Ala Gln Glu Gly
115 120 125
Asn Ile Gln Asn Thr Val Met Leu Asp Lys Gln Lys Glu Leu Asp Ser
130 135 140
Lys Val Arg Asn Val Lys Asp Gln Val Met Cys Ile Glu Gln Glu Ile
145 150 155 160
Lys Thr Leu Glu Glu Leu Gln Asp Glu Tyr Asp Phe Lys Cys Lys Thr
165 170 175
Ser Gln Asn Arg Glu Gly Glu Ala Asn Gly Val Ala Lys Ser Asp Gln
180 185 190
Lys Gln Glu Gln Leu Leu Leu His Lys Met Phe Leu Met Leu Asp Asn
195 200 205
Lys Arg Lys Glu Ile Ile His Lys Ile Arg Glu Leu Leu Asn Ser Ile
210 215 220
Glu Leu Thr Gln Asn Thr Leu Ile Asn Asp Glu Leu Val Glu Trp Lys
225 230 235 240
Arg Arg Gln Gln Ser Ala Cys Ile Gly Gly Pro Pro Asn Ala Cys Leu
245 250 255
Asp Gln Leu Gln Thr Trp Phe Thr Ile Val Ala Glu Thr Leu Gln Gln
260 265 270

Ile Arg Gln Gln Leu Lys Lys Leu Glu Glu Leu Glu Gln Lys Phe Thr
 275 280 285
 Tyr Glu Pro Asp Pro Ile Thr Lys Asn Lys Gln Val Leu Ser Asp Arg
 290 295 300
 Thr Phe Leu Leu Phe Gln Gln Leu Ile Gln Ser Ser Phe Val Val Glu
 305 310 315 320
 Arg Gln Pro Cys Met Pro Thr His Pro Gln Arg Pro Leu Val Leu Lys
 325 330 335
 Thr Gly Val Gln Phe Thr Val Lys Ser Arg Leu Leu Val Lys Leu Gln
 340 345 350
 Glu Ser Asn Leu Leu Thr Lys Val Lys Cys His Phe Asp Lys Asp Val
 355 360 365
 Asn Glu Lys Asn Thr Val Lys Gly Phe Arg Lys Phe Asn Ile Leu Gly
 370 375 380
 Thr His Thr Lys Val Met Asn Met Glu Glu Ser Thr Asn Gly Ser Leu
 385 390 395 400
 Ala Ala Glu Leu Arg His Leu Gln Leu Lys Glu Gln Lys Asn Ala Gly
 405 410 415
 Asn Arg Thr Asn Glu Gly Pro Leu Ile Val Thr Glu Glu Leu His Ser
 420 425 430
 Leu Ser Phe Glu Thr Gln Leu Cys Gln Pro Gly Leu Val Ile Asp Leu
 435 440 445
 Glu Thr Thr Ser Leu Pro Val Val Val Ile Ser Asn Val Ser Gln Leu
 450 455 460
 Pro Ser Gly Trp Ala Ser Ile Leu Trp Tyr Asn Met Leu Val Thr Glu
 465 470 475 480
 Pro Arg Asn Leu Ser Phe Phe Leu Asn Pro Pro Cys Ala Trp Trp Ser
 485 490 495
 Gln Leu Ser Glu Val Leu Ser Trp Gln Phe Ser Ser Val Thr Lys Arg
 500 505 510
 Gly Leu Asn Ala Asp Gln Leu Ser Met Leu Gly Glu Lys Leu Leu Gly
 515 520 525
 Pro Asn Ala Gly Pro Asp Gly Leu Ile Pro Trp Thr Arg Phe Cys Lys
 530 535 540
 Glu Asn Ile Asn Asp Lys Asn Phe Ser Phe Trp Pro Trp Ile Asp Thr
 545 550 555 560
 Ile Leu Glu Leu Ile Lys Asn Asp Leu Leu Cys Leu Trp Asn Asp Gly
 565 570 575
 Cys Ile Met Gly Phe Ile Ser Lys Glu Arg Glu Arg Ala Leu Leu Lys
 580 585 590
 Asp Gln Gln Pro Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser Ser Arg
 595 600 605
 Glu Gly Ala Ile Thr Phe Thr Trp Val Glu Arg Ser Gln Asn Gly Gly
 610 615 620
 Glu Pro Asp Phe His Ala Val Glu Pro Tyr Thr Lys Lys Glu Leu Ser
 625 630 635 640

Ala Val Thr Phe Pro Asp Ile Ile Arg Asn Tyr Lys Val Met Ala Ala
645 650 655

Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro Asn Ile Asp
660 665 670

Lys Asp His Ala Phe Gly Lys Tyr Tyr Ser Arg Pro Lys Glu Ala Pro
675 680 685

Glu Pro Met Glu Leu Asp Asp Pro Lys Arg Thr Gly Tyr Ile Lys Thr
690 695 700

Glu Leu Ile Ser Val Ser Glu Val His Pro Ser Arg Leu Gln Thr Thr
705 710 715 720

Asp Asn Leu Leu Pro Met Ser Pro Glu Glu Phe Asp Glu Met Ser Arg
725 730 735

Ile Val Gly Pro Glu Phe Asp Ser Met Met Ser Thr Val
740 745

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2375 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: splenic/thymic
(B) CLONE: Murine 13sf1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 34..2277

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGCCACTACC TGGACGGAGA GAGAGAGAGC AGC ATG TCT CAG TGG AAT CAA GTC	54
Met Ser Gln Trp Asn Gln Val	
1 5	
CAA CAA TTA GAA ATC AAG TTT TTG GAG CAA GTA GAT CAG TTC TAT GAT	102
Gln Gln Leu Glu Ile Lys Phe Leu Glu Gln Val Asp Gln Phe Tyr Asp	
10 15 20	
GAC AAC TTT CCT ATG GAA ATC CGG CAT CTG CTA GCT CAG TGG ATT GAG	150
Asp Asn Phe Pro Met Glu Ile Arg His Leu Leu Ala Gln Trp Ile Glu	
25 30 35	
ACT CAA GAC TGG GAA GTA GCT TCT AAC AAT GAA ACT ATG GCA ACA ATT	198
Thr Gln Asp Trp Glu Val Ala Ser Asn Asn Glu Thr Met Ala Thr Ile	
40 45 50 55	
CTG CTT CAA AAC TTA CTA ATA CAA TTG GAT GAA CAG TTG GGG CGG GTT	246
Leu Leu Gln Asn Leu Leu Ile Gln Leu Asp Glu Gln Leu Gly Arg Val	
60 65 70	

TCC	AAA	GAA	AAA	AAT	CTG	CTA	TTG	ATT	CAC	AAT	CTA	AAG	AGA	ATT	AGA	294
Ser	Lys	Glu	Lys	Asn	Leu	Leu	Leu	Ile	His	Asn	Leu	Lys	Arg	Ile	Arg	
			75					80					85			
AAA	GTT	CTT	CAG	GGC	AAG	TTT	CAT	GGA	AAT	CCA	ATG	CAT	GTA	GCT	GTG	342
Lys	Val	Leu	Gln	Gly	Lys	Phe	His	Gly	Asn	Pro	Met	His	Val	Ala	Val	
		90					95					100				
GTA	ATT	TCA	AAT	TGC	TTA	AGG	GAA	GAG	AGG	AGA	ATA	TTG	GCT	GCA	GCC	390
Val	Ile	Ser	Asn	Cys	Leu	Arg	Glu	Glu	Arg	Arg	Ile	Leu	Ala	Ala	Ala	
	105					110					115					
AAC	ATG	CCT	ATC	CAG	GGA	CCT	CTG	GAG	AAA	TCC	TTA	CAG	AGT	TCT	TCA	438
Asn	Met	Pro	Ile	Gln	Gly	Pro	Leu	Glu	Lys	Ser	Leu	Gln	Ser	Ser	Ser	
120					125					130					135	
GTT	TCT	GAA	AGA	CAA	AGG	AAT	GTG	GAA	CAC	AAA	GTG	TCT	GCC	ATT	AAA	486
Val	Ser	Glu	Arg	Gln	Arg	Asn	Val	Glu	His	Lys	Val	Ser	Ala	Ile	Lys	
			140						145					150		
AAC	AGT	GTG	CAG	ATG	ACA	GAA	CAA	GAT	ACC	AAA	TAC	TTA	GAA	GAC	CTG	534
Asn	Ser	Val	Gln	Met	Thr	Glu	Gln	Asp	Thr	Lys	Tyr	Leu	Glu	Asp	Leu	
			155					160					165			
CAA	GAT	GAG	TTT	GAC	TAC	AGG	TAT	AAA	ACA	ATT	CAG	ACA	ATG	GAT	CAG	582
Gln	Asp	Glu	Phe	Asp	Tyr	Arg	Tyr	Lys	Thr	Ile	Gln	Thr	Met	Asp	Gln	
		170					175					180				
GGT	GAC	AAA	AAC	AGT	ATC	CTG	GTG	AAC	CAG	GAA	GTT	TTG	ACA	CTG	CTG	630
Gly	Asp	Lys	Asn	Ser	Ile	Leu	Val	Asn	Gln	Glu	Val	Leu	Thr	Leu	Leu	
	185					190					195					
CAA	GAA	ATG	CTT	AAT	AGT	CTG	GAC	TTC	AAG	AGA	AAG	GAA	GCA	CTC	AGT	678
Gln	Glu	Met	Leu	Asn	Ser	Leu	Asp	Phe	Lys	Arg	Lys	Glu	Ala	Leu	Ser	
200					205					210					215	
AAG	ATG	ACG	CAG	ATA	GTG	AAC	GAG	ACA	GAC	CTG	CTC	ATG	AAC	AGC	ATG	726
Lys	Met	Thr	Gln	Ile	Val	Asn	Glu	Thr	Asp	Leu	Leu	Met	Asn	Ser	Met	
				220					225					230		
CTT	CTA	GAA	GAG	CTG	CAG	GAC	TGG	AAA	AAG	CGG	CAC	AGG	ATT	GCC	TGC	774
Leu	Leu	Glu	Gln	Leu	Gln	Asp	Trp	Lys	Lys	Arg	His	Arg	Ile	Ala	Cys	
			235					240					245			
ATT	GGT	GGC	CCG	CTC	CAC	AAT	GGG	CTG	GAC	CAG	CTT	CAG	AAC	TGC	TTT	822
Ile	Gly	Gly	Pro	Leu	His	Asn	Gly	Leu	Asp	Gln	Leu	Gln	Asn	Cys	Phe	
		250					255					260				
ACC	CTA	CTG	GCA	GAG	AGT	CTT	TTC	CAA	CTC	AGA	CAG	CAA	CTG	GAG	AAA	870
Thr	Leu	Leu	Ala	Glu	Ser	Leu	Phe	Gln	Leu	Arg		Gln	Leu	Glu	Lys	
	265					270				275						
CTA	CAG	GAG	CAA	TCT	ACT	AAA	ATG	ACC	TAT	GAA	GGG	GAT	CCC	ATC	CCT	918
Leu	Gln	Glu	Gln	Ser	Thr	Lys	Met	Thr	Tyr	Glu	Gly	Asp	Pro	Ile	Pro	
280					285					290					295	
GCT	CAA	AGA	GCA	CAC	CTC	CTG	GAA	AGA	GCT	ACC	TTC	CTG	ATC	TAC	AAC	966
Ala	Gln	Arg	Ala	His	Leu	Leu	Glu	Arg	Ala	Thr	Phe	Leu	Ile	Tyr	Asn	
				300					305					310		
CTT	TTC	AAG	AAC	TCA	TTT	GTG	GTC	GAG	CGA	CAC	GCA	TGC	ATG	CCA	ACG	1014
Leu	Phe	Lys	Asn	Ser	Phe	Val	Val	Glu	Arg	His	Ala	Cys	Met	Pro	Thr	
			315					320					325			
CAC	CCT	CAG	AGG	CCG	ATG	GTA	CTT	AAA	ACC	CTC	ATT	CAG	TTC	ACT	GTA	1062
His	Pro	Gln	Arg	Pro	Met	Val	Leu	Lys	Thr	Leu	Ile	Gln	Phe	Thr	Val	
		330					335					340				

AAA CTG AGA TTA CTA ATA AAA TTG CCG GAA CTA AAC TAT CAG GTG AAA	1110
Lys Leu Arg Leu Leu Ile Lys Leu Pro Glu Leu Asn Tyr Gln Val Lys	
345 350 355	
GTA AAG GCG TCC ATT GAC AAG AAT GTT TCA ACT CTA AGC AAT AGA AGA	1158
Val Lys Ala Ser Ile Asp Lys Asn Val Ser Thr Leu Ser Asn Arg Arg	
360 365 370 375	
TTT GTG CTT TGT GGA ACT CAC GTC AAA GCT ATG TCC AGT GAG GAA TCT	1206
Phe Val Leu Cys Gly Thr His Val Lys Ala Met Ser Ser Glu Glu Ser	
380 385 390	
TCC AAT GGG AGC CTC TCA GTG GAG TTA GAC ATT GCA ACC CAA GGA GAT	1254
Ser Asn Gly Ser Leu Ser Val Glu Leu Asp Ile Ala Thr Gln Gly Asp	
395 400 405	
GAA GTG CAG TAC TGG AGT AAA GGA AAC GAG GGC TGC CAC ATG GTG ACA	1302
Glu Val Gln Tyr Trp Ser Lys Gly Asn Glu Gly Cys His Met Val Thr	
410 415 420	
GAG GAG TTG CAT TCC ATA ACC TTT GAG ACC CAG ATC TGC CTC TAT GGC	1350
Glu Glu Leu His Ser Ile Thr Phe Glu Thr Gln Ile Cys Leu Tyr Gly	
425 430 435	
CTC ACC ATT AAC CTA GAG ACC AGC TCA TTA CCT GTC GTG ATG ATT TCT	1398
Leu Thr Ile Asn Leu Glu Thr Ser Ser Leu Pro Val Val Met Ile Ser	
440 445 450 455	
AAT GTC AGC CAA CTA CCT AAT GCA TGG GCA TCC ATC ATT TGG TAC AAT	1446
Asn Val Ser Gln Leu Pro Asn Ala Trp Ala Ser Ile Ile Trp Tyr Asn	
460 465 470	
GTA TCA ACT AAC GAC TCC CAG AAC TTG GTT TTC TTT AAT AAC CCT CCA	1494
Val Ser Thr Asn Asp Ser Gln Asn Leu Val Phe Phe Asn Asn Pro Pro	
475 480 485	
TCT GTC ACT TTG GGC CAA CTC CTG GAA GTG ATG AGC TGG CAA TTT TCA	1542
Ser Val Thr Leu Gly Gln Leu Leu Glu Val Met Ser Trp Gln Phe Ser	
490 495 500	
TCC TAT GTC GGT CGT GGC CTT AAT TCA GAG CAG CTC AAC ATG CTG GCA	1590
Ser Tyr Val Gly Arg Gly Leu Asn Ser Glu Gln Leu Asn Met Leu Ala	
505 510 515	
GAG AAG CTC ACA GTT CAG TCT AAC TAC AAT GAT GGT CAC CTC ACC TGG	1638
Glu Lys Leu Thr Val Gln Ser Asn Tyr Asn Asp Gly His Leu Thr Trp	
520 525 530 535	
GCC AAG TTC TGC AAG GAA CAT TTG CCT GGC AAA ACA TTT ACC TTC TGG	1686
Ala Lys Phe Cys Lys Glu His Leu Pro Gly Lys Thr Phe Thr Phe Trp	
540 545 550	
ACT TGG CTT GAA GCA ATA TTG GAC CTA ATT AAA AAA CAT ATT CTT CCC	1734
Thr Trp Leu Glu Ala Ile Leu Asp Leu Ile Lys Lys His Ile Leu Pro	
555 560 565	
CTC TGG ATT GAT GGG TAC ATC ATG GGA TTT GTT AGT AAA GAG AAG GAA	1782
Leu Trp Ile Asp Gly Tyr Ile Met Gly Phe Val Ser Lys Glu Lys Glu	
570 575 580	
CGG CTT CTG CTC AAA GAT AAA ATG CCT GGG ACA TTT TTG TTA AGA TTC	1830
Arg Leu Leu Leu Lys Asp Lys Met Pro Gly Thr Phe Leu Leu Arg Phe	
585 590 595	
AGT GAG AGC CAT CTT GGA GGG ATA ACC TTC ACC TGG GTG GAC CAA TCT	1878
Ser Glu Ser His Leu Gly Gly Ile Thr Phe Thr Trp Val Asp Gln Ser	
600 605 610 615	

GAA AAT GGA GAA GTG AGA TTC CAC TCT GTA GAA CCC TAC AAC AAA GGG Glu Asn Gly Glu Val Arg Phe His Ser Val Glu Pro Tyr Asn Lys Gly 620 625 630	1926
AGA CTG TCG GCT CTG GCC TTC GCT GAC ATC CTG CGA GAC TAC AAG GTT Arg Leu Ser Ala Leu Ala Phe Ala Asp Ile Leu Arg Asp Tyr Lys Val 635 640 645	1974
ATC ATG GCT GAA AAC ATC CCT GAA AAC CCT CTG AAG TAC CTC TAC CCT Ile Met Ala Glu Asn Ile Pro Glu Asn Pro Leu Lys Tyr Leu Tyr Pro 650 655 660	2022
GAC ATT CCC AAA GAC AAA GCC TTT GGC AAA CAC TAC AGC TCC CAG CCG Asp Ile Pro Lys Asp Lys Ala Phe Gly Lys His Tyr Ser Ser Gln Pro 665 670 675	2070
TGC GAA GTC TCA AGA CCA ACC GAA CGG GGA GAC AAG GGT TAC GTC CCC Cys Glu Val Ser Arg Pro Thr Glu Arg Gly Asp Lys Gly Tyr Val Pro 680 685 690 695	2118
TCT GTT TTT ATC CCC ATT TCA ACA ATC CGA AGC GAT TCC ACG GAG CCA Ser Val Phe Ile Pro Ile Ser Thr Ile Arg Ser Asp Ser Thr Glu Pro 700 705 710	2166
CAA TCT CCT TCA GAC CTT CTC CCC ATG TCT CCA AGT GCA TAT GCT GTG Gln Ser Pro Ser Asp Leu Leu Pro Met Ser Pro Ser Ala Tyr Ala Val 715 720 725	2214
CTG AGA GAA AAC CTG AGC CCA ACG ACA ATT GAA ACT GCA ATG AAT TCC Leu Arg Glu Asn Leu Ser Pro Thr Thr Ile Glu Thr Ala Met Asn Ser 730 735 740	2262
CCA TAT TCT GCT GAA TGACGGTGCA AACGGACACT TTAAAGAAGG AAGCAGATGA Pro Tyr Ser Ala Glu 745	2317
AACTGGAGAG TGTTCTTTAC CATAGATCAC AATTTATTTTC TTCGGCTTTG TAAATACC	2375

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 748 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Ser Gln Trp Asn Gln Val Gln Gln Leu Glu Ile Lys Phe Leu Glu 1 5 10 15
Gln Val Asp Gln Phe Tyr Asp Asp Asn Phe Pro Met Glu Ile Arg His 20 25 30
Leu Leu Ala Gln Trp Ile Glu Thr Gln Asp Trp Glu Val Ala Ser Asn 35 40 45
Asn Glu Thr Met Ala Thr Ile Leu Leu Gln Asn Leu Leu Ile Gln Leu 50 55 60
Asp Glu Gln Leu Gly Arg Val Ser Lys Glu Lys Asn Leu Leu Leu Ile 65 70 75 80
His Asn Leu Lys Arg Ile Arg Lys Val Leu Gln Gly Lys Phe His Gly 85 90 95

Asn Pro Met His Val Ala Val Val Ile Ser Asn Cys Leu Arg Glu Glu
 100 105 110
 Arg Arg Ile Leu Ala Ala Ala Asn Met Pro Ile Gln Gly Pro Leu Glu
 115 120 125
 Lys Ser Leu Gln Ser Ser Ser Val Ser Glu Arg Gln Arg Asn Val Glu
 130 135 140
 His Lys Val Ser Ala Ile Lys Asn Ser Val Gln Met Thr Glu Gln Asp
 145 150 155 160
 Thr Lys Tyr Leu Glu Asp Leu Gln Asp Glu Phe Asp Tyr Arg Tyr Lys
 165 170 175
 Thr Ile Gln Thr Met Asp Gln Gly Asp Lys Asn Ser Ile Leu Val Asn
 180 185 190
 Gln Glu Val Leu Thr Leu Leu Gln Glu Met Leu Asn Ser Leu Asp Phe
 195 200 205
 Lys Arg Lys Glu Ala Leu Ser Lys Met Thr Gln Ile Val Asn Glu Thr
 210 215 220
 Asp Leu Leu Met Asn Ser Met Leu Leu Glu Glu Leu Gln Asp Trp Lys
 225 230 235 240
 Lys Arg His Arg Ile Ala Cys Ile Gly Gly Pro Leu His Asn Gly Leu
 245 250 255
 Asp Gln Leu Gln Asn Cys Phe Thr Leu Leu Ala Glu Ser Leu Phe Gln
 260 265 270
 Leu Arg Gln Gln Leu Glu Lys Leu Gln Glu Gln Ser Thr Lys Met Thr
 275 280 285
 Tyr Glu Gly Asp Pro Ile Pro Ala Gln Arg Ala His Leu Leu Glu Arg
 290 295 300
 Ala Thr Phe Leu Ile Tyr Asn Leu Phe Lys Asn Ser Phe Val Val Glu
 305 310 315 320
 Arg His Ala Cys Met Pro Thr His Pro Gln Arg Pro Met Val Leu Lys
 325 330 335
 Thr Leu Ile Gln Phe Thr Val Lys Leu Arg Leu Leu Ile Lys Leu Pro
 340 345 350
 Glu Leu Asn Tyr Gln Val Lys Val Lys Ala Ser Ile Asp Lys Asn Val
 355 360 365
 Ser Thr Leu Ser Asn Arg Arg Phe Val Leu Cys Gly Thr His Val Lys
 370 375 380
 Ala Met Ser Ser Glu Glu Ser Ser Asn Gly Ser Leu Ser Val Glu Leu
 385 390 395 400
 Asp Ile Ala Thr Gln Gly Asp Glu Val Gln Tyr Trp Ser Lys Gly Asn
 405 410 415
 Glu Gly Cys His Met Val Thr Glu Glu Leu His Ser Ile Thr Phe Glu
 420 425 430
 Thr Gln Ile Cys Leu Tyr Gly Leu Thr Ile Asn Leu Glu Thr Ser Ser
 435 440 445
 Leu Pro Val Val Met Ile Ser Asn Val Ser Gln Leu Pro Asn Ala Trp
 450 455 460

Ala Ser Ile Ile Trp Tyr Asn Val Ser Thr Asn Asp Ser Gln Asn Leu
 465 470 475 480
 Val Phe Phe Asn Asn Pro Pro Ser Val Thr Leu Gly Gln Leu Leu Glu
 485 490 495
 Val Met Ser Trp Gln Phe Ser Ser Tyr Val Gly Arg Gly Leu Asn Ser
 500 505 510
 Glu Gln Leu Asn Met Leu Ala Glu Lys Leu Thr Val Gln Ser Asn Tyr
 515 520 525
 Asn Asp Gly His Leu Thr Trp Ala Lys Phe Cys Lys Glu His Leu Pro
 530 535 540
 Gly Lys Thr Phe Thr Phe Trp Thr Trp Leu Glu Ala Ile Leu Asp Leu
 545 550 555 560
 Ile Lys Lys His Ile Leu Pro Leu Trp Ile Asp Gly Tyr Ile Met Gly
 565 570 575
 Phe Val Ser Lys Glu Lys Glu Arg Leu Leu Lys Asp Lys Met Pro
 580 585 590
 Gly Thr Phe Leu Leu Arg Phe Ser Glu Ser His Leu Gly Gly Ile Thr
 595 600 605
 Phe Thr Trp Val Asp Gln Ser Glu Asn Gly Glu Val Arg Phe His Ser
 610 615 620
 Val Glu Pro Tyr Asn Lys Gly Arg Leu Ser Ala Leu Ala Phe Ala Asp
 625 630 635 640
 Ile Leu Arg Asp Tyr Lys Val Ile Met Ala Glu Asn Ile Pro Glu Asn
 645 650 655
 Pro Leu Lys Tyr Leu Tyr Pro Asp Ile Pro Lys Asp Lys Ala Phe Gly
 660 665 670
 Lys His Tyr Ser Ser Gln Pro Cys Glu Val Ser Arg Pro Thr Glu Arg
 675 680 685
 Gly Asp Lys Gly Tyr Val Pro Ser Val Phe Ile Pro Ile Ser Thr Ile
 690 695 700
 Arg Ser Asp Ser Thr Glu Pro Gln Ser Pro Ser Asp Leu Leu Pro Met
 705 710 715 720
 Ser Pro Ser Ala Tyr Ala Val Leu Arg Glu Asn Leu Ser Pro Thr Thr
 725 730 735
 Ile Glu Thr Ala Met Asn Ser Pro Tyr Ser Ala Glu
 740 745

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2869 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mouse

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: splenic/thymic

(B) CLONE: Murine.19sf6

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 69..2378

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCCGCGACCA GCCAGGCCGG CCAGTCGGGC TCAGCCCGGA GACAGTCGAG ACCCCTGACT	60
GCAGCAGG ATG GCT CAG TGG AAC CAG CTG CAG CAG CTG GAC ACA CGC TAC	110
Met Ala Gln Trp Asn Gln Leu Gln Gln Leu Asp Thr Arg Tyr	
1 5 10	
CTG AAG CAG CTG CAC CAG CTG TAC AGC GAC ACG TTC CCC ATG GAG CTG	158
Leu Lys Gln Leu His Gln Leu Tyr Ser Asp Thr Phe Pro Met Glu Leu	
15 20 25 30	
CGG CAG TTC CTG GCA CCT TGG ATT GAG AGT CAA GAC TGG GCA TAT GCA	206
Arg Gln Phe Leu Ala Pro Trp Ile Glu Ser Gln Asp Trp Ala Tyr Ala	
35 40 45	
GCC AGC AAA GAG TCA CAT GCC ACG TTG GTG TTT CAT AAT CTC TTG GGT	254
Ala Ser Lys Glu Ser His Ala Thr Leu Val Phe His Asn Leu Leu Gly	
50 55 60	
GAA ATT GAC CAG CAA TAT AGC CGA TTC CTG CAA GAG TCC AAT GTC CTC	302
Glu Ile Asp Gln Gln Tyr Ser Arg Phe Leu Gln Glu Ser Asn Val Leu	
65 70 75	
TAT CAG CAC AAC CTT CGA AGA ATC AAG CAG TTT CTG CAG AGC AGG TAT	350
Tyr Gln His Asn Leu Arg Arg Ile Lys Gln Phe Leu Gln Ser Arg Tyr	
80 85 90	
CTT GAG AAG CCA ATG GAA ATT GCC CGG ATC GTG GCC CGA TGC CTG TGG	398
Leu Glu Lys Pro Met Glu Ile Ala Arg Ile Val Ala Arg Cys Leu Trp	
95 100 105 110	
GAA GAG TCT CGC CTC CTC CAG ACG GCA GCC ACG GCA GCC CAG CAA GGG	446
Glu Glu Ser Arg Leu Leu Gln Thr Ala Ala Thr Ala Ala Gln Gln Gly	
115 120 125	
GGC CAG GCC AAC CAC CCA ACA GCC GCC GTA GTG ACA GAG AAG CAG CAG	494
Gly Gln Ala Asn His Pro Thr Ala Ala Val Val Thr Glu Lys Gln Gln	
130 135 140	
ATG TTG GAG CAG CAT CTT CAG GAT GTC CGG AAG CGA GTG CAG GAT CTA	542
Met Leu Glu Gln His Leu Gln Asp Val Arg Lys Arg Val Gln Asp Leu	
145 150 155	
GAA CAG AAA ATG AAG GTG GTG GAG AAC CTC CAG GAC GAC TTT GAT TTC	590
Glu Gln Lys Met Lys Val Val Glu Asn Leu Gln Asp Asp Phe Asp Phe	
160 165 170	
AAC TAC AAA ACC CTC AAG AGC CAA GGA GAC ATG CAG GAT CTG AAT GGA	638
Asn Tyr Lys Thr Leu Lys Ser Gln Gly Asp Met Gln Asp Leu Asn Gly	
175 180 185 190	
AAC AAC CAG TCT GTG ACC AGA CAG AAG ATG CAG CAG CTG GAA CAG ATG	686
Asn Asn Gln Ser Val Thr Arg Gln Lys Met Gln Gln Leu Glu Gln Met	
195 200 205	

CTC	ACA	GCC	CTG	GAC	CAG	ATG	CGG	AGA	AGC	ATT	GTG	AGT	GAG	CTG	GCG	734
Leu	Thr	Ala	Leu	Asp	Gln	Met	Arg	Arg	Ser	Ile	Val	Ser	Glu	Leu	Ala	
			210					215					220			
GGG	CTC	TTG	TCA	GCA	ATG	GAG	TAC	GTG	CAG	AAG	ACA	CTG	ACT	GAT	GAA	782
Gly	Leu	Leu	Ser	Ala	Met	Glu	Tyr	Val	Gln	Lys	Thr	Leu	Thr	Asp	Glu	
		225					230					235				
GAG	CTG	GCT	GAC	TGG	AAG	AGG	CGG	CCA	GAG	ATC	GCG	TGC	ATC	GGA	GGC	830
Glu	Leu	Ala	Asp	Trp	Lys	Arg	Arg	Pro	Glu	Ile	Ala	Cys	Ile	Gly	Gly	
	240					245					250					
CCT	CCC	AAC	ATC	TGC	CTG	GAC	CGT	CTG	GAA	AAC	TGG	ATA	ACT	TCA	TTA	878
Pro	Pro	Asn	Ile	Cys	Leu	Asp	Arg	Leu	Glu	Asn	Trp	Ile	Thr	Ser	Leu	
255					260					265					270	
GCA	GAA	TCT	CAA	CTT	CAG	ACC	CGC	CAA	CAA	ATT	AAG	AAA	CTG	GAG	GAG	926
Ala	Glu	Ser	Gln	Leu	Gln	Thr	Arg	Gln	Gln	Ile	Lys	Lys	Leu	Glu	Glu	
				275					280					285		
CTG	CAG	CAG	AAA	GTG	TCC	TAC	AAG	GGC	GAC	CCT	ATC	GTG	CAG	CAC	CGG	974
Leu	Gln	Gln	Lys	Val	Ser	Tyr	Lys	Gly	Asp	Pro	Ile	Val	Gln	His	Arg	
			290					295					300			
CCC	ATG	CTG	GAG	GAG	AGG	ATC	GTG	GAG	CTG	TTC	AGA	AAC	TTA	ATG	AAG	1022
Pro	Met	Leu	Glu	Glu	Arg	Ile	Val	Glu	Leu	Phe	Arg	Asn	Leu	Met	Lys	
		305					310					315				
AGT	GCC	TTC	GTG	GTG	GAG	CGG	CAG	CCC	TGC	ATG	CCC	ATG	CAC	CCG	GAC	1070
Ser	Ala	Phe	Val	Val	Glu	Arg	Gln	Pro	Cys	Met	Pro	Met	His	Pro	Asp	
	320					325					330					
CGG	CCC	TTA	GTC	ATC	AAG	ACT	GGT	GTC	CAG	TTT	ACC	ACG	AAA	GTC	AGG	1118
Arg	Pro	Leu	Val	Ile	Lys	Thr	Gly	Val	Gln	Phe	Thr	Thr	Lys	Val	Arg	
335					340					345					350	
TTG	CTG	GTC	AAA	TTT	CCT	GAG	TTG	AAT	TAT	CAG	CTT	AAA	ATT	AAA	GTG	1166
Leu	Leu	Val	Lys	Phe	Pro	Glu	Leu	Asn	Tyr	Gln	Leu	Lys	Ile	Lys	Val	
				355					360					365		
TGC	ATT	GAT	AAA	GAC	TCT	GGG	GAT	GTT	GCT	GCC	CTC	AGA	GGG	TCT	CGG	1214
Cys	Ile	Asp	Lys	Asp	Ser	Gly	Asp	Val	Ala	Ala	Leu	Arg	Gly	Ser	Arg	
			370					375					380			
AAA	TTT	AAC	ATT	CTG	GGC	ACG	AAC	ACA	AAA	GTG	ATG	AAC	ATG	GAG	GAG	1262
Lys	Phe	Asn	Ile	Leu	Gly	Thr	Asn	Thr	Lys	Val	Met	Asn	Met	Glu	Glu	
		385					390					395				
TCT	AAC	AAC	GGC	AGC	CTG	TCT	GCA	GAG	TTC	AAG	CAC	CTG	ACC	CTT	AGG	1310
Ser	Asn	Asn	Gly	Ser	Leu	Ser	Ala	Glu	Phe	Lys	His	Leu	Thr	Leu	Arg	
	400					405					410					
GAG	CAG	AGA	TGT	GGG	AAT	GGA	GGC	CGT	GCC	AAT	TGT	GAT	GCC	TCC	TTG	1358
Glu	Gln	Arg	Cys	Gly	Asn	Gly	Gly	Arg	Ala	Asn	Cys	Asp	Ala	Ser	Leu	
415					420					425					430	
ATC	GTG	ACT	GAG	GAG	CTG	CAC	CTG	ATC	ACC	TTC	GAG	ACT	GAG	GTG	TAC	1406
Ile	Val	Thr	Glu	Glu	Leu	His	Leu	Ile	Thr	Phe	Glu	Thr	Glu	Val	Tyr	
				435					440					445		
CAC	CAA	GGC	CTC	AAG	ATT	GAC	CTA	GAG	ACC	CAC	TCC	TTG	CCA	GTT	GTG	1454
His	Gln	Gly	Leu	Lys	Ile	Asp	Leu	Glu	Thr	His	Ser	Leu	Pro	Val	Val	
			450					455					460			
GTG	ATC	TCC	AAC	ATC	TGT	CAG	ATG	CCA	AAT	GCT	TGG	GCA	TCA	ATC	CTG	1502
Val	Ile	Ser	Asn	Ile	Cys	Gln	Met	Pro	Asn	Ala	Trp	Ala	Ser	Ile	Leu	
		465					470					475				

TGG Trp	TAT Tyr	AAC Asn	ATG Met	CTG Leu	ACC Thr	AAT Asn	AAC Asn	CCC Pro	AAG Lys	AAC Asn	GTG Val	AAC Asn	TTC Phe	TTC Phe	ACT Thr	1550
480						485					490					
AAG Lys	CCG Pro	CCA Pro	ATT Ile	GGA Gly	ACC Thr	TGG Trp	GAC Asp	CAA Gln	GTG Val	GCC Ala	GAG Glu	GTG Val	CTC Leu	AGC Ser	TGG Trp	1598
495					500					505					510	
CAG Gln	TTC Phe	TCG Ser	TCC Ser	ACC Thr	ACC Thr	AAG Lys	CGA Arg	GGG Gly	CTG Leu	AGC Ser	ATC Ile	GAG Glu	CAG Gln	CTG Leu	ACA Thr	1646
				515					520					525		
ACG Thr	CTG Leu	GCT Ala	GAG Glu	AAG Lys	CTC Leu	CTA Leu	GGG Gly	CCT Pro	GGT Gly	GTG Val	AAC Asn	TAC Tyr	TCA Ser	GGG Gly	TGT Cys	1694
			530					535					540			
CAG Gln	ATC Ile	ACA Thr	TGG Trp	GCT Ala	AAA Lys	TTC Phe	TGC Cys	AAA Lys	GAA Glu	AAC Asn	ATG Met	GCT Ala	GGC Gly	AAG Lys	GGC Gly	1742
		545					550					555				
TTC Phe	TCC Ser	TTC Phe	TGG Trp	GTC Val	TGG Trp	CTA Leu	GAC Asp	AAT Asn	ATC Ile	ATC Ile	GAC Asp	CTT Leu	GTG Val	AAA Lys	AAG Lys	1790
	560					565					570					
TAT Tyr	ATC Ile	TTG Leu	GCC Ala	CTT Leu	TGG Trp	AAT Asn	GAA Glu	GGG Gly	TAC Tyr	ATC Ile	ATG Met	GGT Gly	TTC Phe	ATC Ile	AGC Ser	1838
575					580					585					590	
AAG Lys	GAG Glu	CGG Arg	GAG Glu	CGG Arg	GCC Ala	ATC Ile	CTA Leu	AGC Ser	ACA Thr	AAG Lys	CCC Pro	CCG Pro	GGC Gly	ACC Thr	TTC Phe	1886
				595					600					605		
CTA Leu	CTG Leu	CGC Arg	TTC Phe	AGC Ser	GAG Glu	AGC Ser	AGC Ser	AAA Lys	GAA Glu	GGA Gly	GGG Gly	GTC Val	ACT Thr	TTC Phe	ACT Thr	1934
			610					615					620			
TGG Trp	GTG Val	GAA Glu	AAG Lys	GAC Asp	ATC Ile	AGT Ser	GGC Gly	AAG Lys	ACC Thr	CAG Gln	ATC Ile	CAG Gln	TCT Ser	GTA Val	GAG Glu	1982
		625					630					635				
CCA Pro	TAC Tyr	ACC Thr	AAG Lys	CAG Gln	CAG Gln	CTG Leu	AAC Asn	AAC Asn	ATG Met	TCA Ser	TTT Phe	GCT Ala	GAA Glu	ATC Ile	ATC Ile	2030
	640					645					650					
ATG Met	GGC Gly	TAT Tyr	AAG Lys	ATC Ile	ATG Met	GAT Asp	GCG Ala	ACC Thr	AAC Asn	ATC Ile	CTG Leu	GTG Val	TCT Ser	CCA Pro	CTT Leu	2078
655					660					665					670	
GTC Val	TAC Tyr	CTC Leu	TAC Tyr	CCC Pro	GAC Asp	ATT Ile	CCC Pro	AAG Lys	GAG Glu	GAG Glu	GCA Ala	TTT Phe	GGA Gly	AAG Lys	TAC Tyr	2126
				675					680				685			
TGT Cys	AGG Arg	CCC Pro	GAG Glu	AGC Ser	CAG Gln	GAG Glu	CAC His	CCC Pro	GAA Glu	GCC Ala	GAC Asp	CCA Pro	GGT Gly	AGT Ser	GCT Ala	2174
			690					695				700				
GCC Ala	CCG Pro	TAC Tyr	CTG Leu	AAG Lys	ACC Thr	AAG Lys	TTC Phe	ATC Ile	TGT Cys	GTG Val	ACA Thr	CCA Pro	ACG Thr	ACC Thr	TGC Cys	2222
		705					710					715				
AGC Ser	AAT Asn	ACC Thr	ATT Ile	GAC Asp	CTG Leu	CCG Pro	ATG Met	TCC Ser	CCC Pro	CGC Arg	ACT Thr	TTA Leu	GAT Asp	TCA Ser	TTG Leu	2270
	720					725					730					
ATG Met	CAG Gln	TTT Phe	GGA Gly	AAT Asn	AAC Asn	GGT Gly	GAA Glu	GGT Gly	GCT Ala	GAG Glu	CCC Pro	TCA Ser	GCA Ala	GGA Gly	GGG Gly	2318
735					740					745					750	

CAG TTT GAG TCG CTC ACG TTT GAC ATG GAT CTG ACC TCG GAG TGT GCT	2366
Gln Phe Glu Ser Leu Thr Phe Asp Met Asp Leu Thr Ser Glu Cys Ala	
755 760 765	
ACC TCC CCC ATG TGAGGAGCTG AAACCAGAAG CTGCAGAGAC GTGACTTGAG	2418
Thr Ser Pro Met	
770	
ACACCTGCCC CGTGCTCCAC CCCTAAGCAG CCGAACCCCA TATCGTCTGA AACTCCTAAC	2478
TTTGTGGTTC CAGATTTTTT TTTTAAATTT CCTACTTCTG CTATCTTTGG GCAATCTGGG	2538
CACTTTTTAA AAGAGAGAAA TGAGTGAGTG TGGGTGATAA ACTGTTATGT AAAGAGGAGA	2598
GACCTCTGAG TCTGGGGATG GGGCTGAGAG CAGAAGGGAG GCAAAGGGGA ACACCTCCTG	2658
TCCTGCCCCG CTGCCCTCCT TTTTCAGCAG CTCGGGGGTT GGTGTGTTAGA CAAGTGCCTC	2718
CTGGTGCCCA TGGCTACCTG TTGCCCCACT CTGTGAGCTG ATACCCCAT TCTGGGAAC TC	2778
CTGGCTCTGC ACTTTCAACC TTGCTAATAT CCACATAGAA GCTAGGACTA AGCCCAGGAG	2838
GTTCTCTTTT AAATTAAAAA AAAAAAAAAA A	2869

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 770 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met	Ala	Gln	Trp	Asn	Gln	Leu	Gln	Gln	Leu	Asp	Thr	Arg	Tyr	Leu	Lys
1				5					10					15	
Gln	Leu	His	Gln	Leu	Tyr	Ser	Asp	Thr	Phe	Pro	Met	Glu	Leu	Arg	Gln
		20					25						30		
Phe	Leu	Ala	Pro	Trp	Ile	Glu	Ser	Gln	Asp	Trp	Ala	Tyr	Ala	Ala	Ser
		35					40					45			
Lys	Glu	Ser	His	Ala	Thr	Leu	Val	Phe	His	Asn	Leu	Leu	Gly	Glu	Ile
	50					55					60				
Asp	Gln	Gln	Tyr	Ser	Arg	Phe	Leu	Gln	Glu	Ser	Asn	Val	Leu	Tyr	Gln
65				70						75				80	
His	Asn	Leu	Arg	Arg	Ile	Lys	Gln	Phe	Leu	Gln	Ser	Arg	Tyr	Leu	Glu
			85						90					95	
Lys	Pro	Met	Glu	Ile	Ala	Arg	Ile	Val	Ala	Arg	Cys	Leu	Trp	Glu	Glu
		100					105						110		
Ser	Arg	Leu	Leu	Gln	Thr	Ala	Ala	Thr	Ala	Ala	Gln	Gln	Gly	Gly	Gln
		115				120						125			
Ala	Asn	His	Pro	Thr	Ala	Ala	Val	Val	Thr	Glu	Lys	Gln	Gln	Met	Leu
	130					135					140				
Glu	Gln	His	Leu	Gln	Asp	Val	Arg	Lys	Arg	Val	Gln	Asp	Leu	Glu	Gln
145				150					155					160	
Lys	Met	Lys	Val	Val	Glu	Asn	Leu	Gln	Asp	Asp	Phe	Asp	Phe	Asn	Tyr
			165						170					175	

Lys Thr Leu Lys Ser Gln Gly Asp Met Gln Asp Leu Asn Gly Asn Asn
 180 185 190
 Gln Ser Val Thr Arg Gln Lys Met Gln Gln Leu Glu Gln Met Leu Thr
 195 200 205
 Ala Leu Asp Gln Met Arg Arg Ser Ile Val Ser Glu Leu Ala Gly Leu
 210 215 220
 Leu Ser Ala Met Glu Tyr Val Gln Lys Thr Leu Thr Asp Glu Glu Leu
 225 230 235 240
 Ala Asp Trp Lys Arg Arg Pro Glu Ile Ala Cys Ile Gly Gly Pro Pro
 245 250 255
 Asn Ile Cys Leu Asp Arg Leu Glu Asn Trp Ile Thr Ser Leu Ala Glu
 260 265 270
 Ser Gln Leu Gln Thr Arg Gln Gln Ile Lys Lys Leu Glu Glu Leu Gln
 275 280 285
 Gln Lys Val Ser Tyr Lys Gly Asp Pro Ile Val Gln His Arg Pro Met
 290 295 300
 Leu Glu Glu Arg Ile Val Glu Leu Phe Arg Asn Leu Met Lys Ser Ala
 305 310 315 320
 Phe Val Val Glu Arg Gln Pro Cys Met Pro Met His Pro Asp Arg Pro
 325 330 335
 Leu Val Ile Lys Thr Gly Val Gln Phe Thr Thr Lys Val Arg Leu Leu
 340 345 350
 Val Lys Phe Pro Glu Leu Asn Tyr Gln Leu Lys Ile Lys Val Cys Ile
 355 360 365
 Asp Lys Asp Ser Gly Asp Val Ala Ala Leu Arg Gly Ser Arg Lys Phe
 370 375 380
 Asn Ile Leu Gly Thr Asn Thr Lys Val Met Asn Met Glu Glu Ser Asn
 385 390 395 400
 Asn Gly Ser Leu Ser Ala Glu Phe Lys His Leu Thr Leu Arg Glu Gln
 405 410 415
 Arg Cys Gly Asn Gly Gly Arg Ala Asn Cys Asp Ala Ser Leu Ile Val
 420 425 430
 Thr Glu Glu Leu His Leu Ile Thr Phe Glu Thr Glu Val Tyr His Gln
 435 440 445
 Gly Leu Lys Ile Asp Leu Glu Thr His Ser Leu Pro Val Val Val Ile
 450 455 460
 Ser Asn Ile Cys Gln Met Pro Asn Ala Trp Ala Ser Ile Leu Trp Tyr
 465 470 475 480
 Asn Met Leu Thr Asn Asn Pro Lys Asn Val Asn Phe Phe Thr Lys Pro
 485 490 495
 Pro Ile Gly Thr Trp Asp Gln Val Ala Glu Val Leu Ser Trp Gln Phe
 500 505 510
 Ser Ser Thr Thr Lys Arg Gly Leu Ser Ile Glu Gln Leu Thr Thr Leu
 515 520 525
 Ala Glu Lys Leu Leu Gly Pro Gly Val Asn Tyr Ser Gly Cys Gln Ile
 530 535 540

Thr Trp Ala Lys Phe Cys Lys Glu Asn Met Ala Gly Lys Gly Phe Ser
 545 550 555 560
 Phe Trp Val Trp Leu Asp Asn Ile Ile Asp Leu Val Lys Lys Tyr Ile
 565 570 575
 Leu Ala Leu Trp Asn Glu Gly Tyr Ile Met Gly Phe Ile Ser Lys Glu
 580 585 590
 Arg Glu Arg Ala Ile Leu Ser Thr Lys Pro Pro Gly Thr Phe Leu Leu
 595 600 605
 Arg Phe Ser Glu Ser Ser Lys Glu Gly Gly Val Thr Phe Thr Trp Val
 610 615 620
 Glu Lys Asp Ile Ser Gly Lys Thr Gln Ile Gln Ser Val Glu Pro Tyr
 625 630 635 640
 Thr Lys Gln Gln Leu Asn Asn Met Ser Phe Ala Glu Ile Ile Met Gly
 645 650 655
 Tyr Lys Ile Met Asp Ala Thr Asn Ile Leu Val Ser Pro Leu Val Tyr
 660 665 670
 Leu Tyr Pro Asp Ile Pro Lys Glu Glu Ala Phe Gly Lys Tyr Cys Arg
 675 680 685
 Pro Glu Ser Gln Glu His Pro Glu Ala Asp Pro Gly Ser Ala Ala Pro
 690 695 700
 Tyr Leu Lys Thr Lys Phe Ile Cys Val Thr Pro Thr Thr Cys Ser Asn
 705 710 715 720
 Thr Ile Asp Leu Pro Met Ser Pro Arg Thr Leu Asp Ser Leu Met Gln
 725 730 735
 Phe Gly Asn Asn Gly Glu Gly Ala Glu Pro Ser Ala Gly Gly Gln Phe
 740 745 750
 Glu Ser Leu Thr Phe Asp Met Asp Leu Thr Ser Glu Cys Ala Thr Ser
 755 760 765
 Pro Met
 770

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAYACNGARC CNATGGARAT YATT

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAYGTNGAYC ARYTNAAYAT G

21

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

RTCDATRTTN GRGTANAR

18

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTAYAANTYR AYCAGNGYAA

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GATCGAGATG TATTTCACG AAAAG

25

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Leu	Asp	Gly	Pro	Lys	Gly	Thr	Gly	Tyr	Ile	Lys	Thr	Glu	Leu	Ile
1				5				10					15	

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Gly	Tyr	Ile	Lys	Thr	Glu
1				5	

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Lys Val Asn Leu Gln Glu Arg Arg Lys Tyr Leu Lys His Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Glu Pro Gln Tyr Glu Glu Ile Pro Ile Tyr Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 105 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vii) IMMEDIATE SOURCE:

(B) CLONE: Src

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Waksman, et al.

(C) JOURNAL: Nature
(D) VOLUME: 358
(F) PAGES: 646-653
(G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ala Glu Glu Trp Tyr Phe Gly Lys Ile Thr Arg Arg Glu Ser Glu Arg
1 5 10 15
Leu Leu Leu Asn Pro Glu Asn Pro Arg Gly Thr Phe Leu Val Arg Glu
20 25 30
Ser Glu Thr Thr Lys Gly Ala Tyr Cys Leu Ser Val Ser Asp Phe Phe
35 40 45
Asp Asn Ala Lys Gly Leu Asn Val Lys His Tyr Lys Ile Arg Lys Leu
50 55 60
Asp Ser Gly Gly Phe Tyr Ile Thr Ser Arg Thr Gln Phe Ser Ser Leu
65 70 75 80
Gln Gln Leu Val Ala Tyr Tyr Ser Lys His Ala Asp Gly Leu Cys His
85 90 95
Arg Leu Thr Asn Val Cys Pro Thr Ser
100 105

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vii) IMMEDIATE SOURCE:

(B) CLONE: Abl

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Overduin, et al.
(C) JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
(D) VOLUME: 89
(F) PAGES: 11673-11677
(G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Glu Lys His Ser Trp Tyr His Gly Pro Val Ser Arg Asn Ala Ala Glu
1 5 10 15
Tyr Leu Leu Ser Ser Gly Ile Asn Gly Ser Phe Leu Val Arg Glu Ser
20 25 30
Asp Arg Arg Pro Gly Gln Arg Ser Ile Ser Leu Arg Tyr Glu Glu Gly
35 40 45
Arg Val Tyr His Tyr Arg Ile Asn Thr Ala Ser Asp Gly Lys Leu Tyr
50 55 60

Val Ser Ser Glu Ser Arg Phe Asn Thr Leu Ala Glu Leu Val His His
65 70 75 80
His Ser Thr Val Ala Asp Gly Leu Ile Thr Thr Leu His Tyr Pro Ala
85 90 95
Pro Lys Arg

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 102 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vii) IMMEDIATE SOURCE:
(B) CLONE: Lck

- (x) PUBLICATION INFORMATION:
(A) AUTHORS: Eck, et al.
(C) JOURNAL: Nature
(D) VOLUME: 362
(F) PAGES: 87-91
(G) DATE: 1993

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Trp Phe Phe Lys Asn Leu Ser Arg Lys Asp Ala Glu Arg Gln Leu Leu
1 5 10 15
Ala Pro Gly Asn Thr His Gly Ser Phe Leu Ile Arg Glu Ser Glu Ser
20 25 30
Thr Ala Gly Ser Phe Ser Leu Ser Val Arg Asp Asp Phe Asp Gln Asn
35 40 45
Gln Gly Glu Val Val Lys His Tyr Lys Ile Arg Asn Leu Asp Asn Gly
50 55 60
Gly Phe Tyr Ile Ser Pro Arg Ile Thr Phe Pro Gly Leu His Asp Leu
65 70 75 80
Val Arg His Tyr Thr Asn Ala Ser Asp Gly Leu Cys Thr Arg Leu Ser
85 90 95
Arg Pro Cys Gln Thr Gln
100

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

[illegible]

姓名	性别	年龄	职业	住址	电话	备注
王德胜	男	45	教师	XX路XX号	XXXX	
李小红	女	38	医生	XX街XX号	XXXX	
张小明	男	25	学生	XX村XX组	XXXX	
赵大刚	男	52	工人	XX厂XX车间	XXXX	
陈丽娟	女	41	护士	XX医院XX科	XXXX	
刘国强	男	33	工程师	XX公司XX部	XXXX	
孙文娟	女	28	会计	XX单位XX处	XXXX	
周志远	男	58	农民	XX乡XX村	XXXX	
吴小芳	女	35	售货员	XX商店XX柜	XXXX	
郑大伟	男	48	干部	XX机关XX科	XXXX	
冯娟娟	女	30	文员	XX单位XX室	XXXX	
马国强	男	55	工人	XX厂XX班	XXXX	
徐小梅	女	42	教师	XX学校XX班	XXXX	
黄大刚	男	36	医生	XX诊所XX室	XXXX	
周丽娟	女	29	学生	XX大学XX系	XXXX	
刘志强	男	50	工人	XX厂XX组	XXXX	
孙文娟	女	43	护士	XX医院XX科	XXXX	
周志远	男	37	工程师	XX公司XX部	XXXX	
吴小芳	女	26	会计	XX单位XX处	XXXX	
郑大伟	男	53	农民	XX乡XX村	XXXX	
冯娟娟	女	34	售货员	XX商店XX柜	XXXX	
马国强	男	47	干部	XX机关XX科	XXXX	
徐小梅	女	31	文员	XX单位XX室	XXXX	
黄大刚	男	56	工人	XX厂XX班	XXXX	
周丽娟	女	44	教师	XX学校XX班	XXXX	
刘志强	男	38	医生	XX诊所XX室	XXXX	
孙文娟	女	27	学生	XX大学XX系	XXXX	
周志远	男	51	工人	XX厂XX组	XXXX	
吴小芳	女	45	护士	XX医院XX科	XXXX	
郑大伟	男	39	工程师	XX公司XX部	XXXX	
冯娟娟	女	29	会计	XX单位XX处	XXXX	
马国强	男	54	农民	XX乡XX村	XXXX	
徐小梅	女	35	售货员	XX商店XX柜	XXXX	
黄大刚	男	49	干部	XX机关XX科	XXXX	
周丽娟	女	32	文员	XX单位XX室	XXXX	
刘志强	男	57	工人	XX厂XX班	XXXX	
孙文娟	女	46	教师	XX学校XX班	XXXX	
周志远	男	40	医生	XX诊所XX室	XXXX	
吴小芳	女	30	学生	XX大学XX系	XXXX	
郑大伟	男	52	工人	XX厂XX组	XXXX	
冯娟娟	女	47	护士	XX医院XX科	XXXX	
马国强	男	41	工程师	XX公司XX部	XXXX	
徐小梅	女	33	会计	XX单位XX处	XXXX	
黄大刚	男	55	农民	XX乡XX村	XXXX	
周丽娟	女	36	售货员	XX商店XX柜	XXXX	
刘志强	男	49	干部	XX机关XX科	XXXX	
孙文娟	女	34	文员	XX单位XX室	XXXX	
周志远	男	58	工人	XX厂XX班	XXXX	
吴小芳	女	48	教师	XX学校XX班	XXXX	
郑大伟	男	42	医生	XX诊所XX室	XXXX	
冯娟娟	女	31	学生	XX大学XX系	XXXX	
马国强	男	53	工人	XX厂XX组	XXXX	
徐小梅	女	49	护士	XX医院XX科	XXXX	
黄大刚	男	43	工程师	XX公司XX部	XXXX	
周丽娟	女	35	会计	XX单位XX处	XXXX	
刘志强	男	56	农民	XX乡XX村	XXXX	
孙文娟	女	37	售货员	XX商店XX柜	XXXX	
周志远	男	50	干部	XX机关XX科	XXXX	
吴小芳	女	38	文员	XX单位XX室	XXXX	
郑大伟	男	59	工人	XX厂XX班	XXXX	
冯娟娟	女	50	教师	XX学校XX班	XXXX	
马国强	男	44	医生	XX诊所XX室	XXXX	
徐小梅	女	32	学生	XX大学XX系	XXXX	
黄大刚	男	54	工人	XX厂XX组	XXXX	
周丽娟	女	51	护士	XX医院XX科	XXXX	
刘志强	男	45	工程师	XX公司XX部	XXXX	
孙文娟	女	39	会计	XX单位XX处	XXXX	
周志远						

PATENT
600-1-195B

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT : JAMES E. DARNELL, JR. ET AL.
SERIAL NO. : UNASSIGNED EXAMINER : UNKNOWN
FILED : HEREWITH ART UNIT : UNKNOWN
FOR : NUCLEIC ACIDS ENCODING RECEPTOR RECOGNITION
FACTORS AND METHODS OF USE THEREOF (AMENDED)

REQUEST FOR TRANSFER OF COMPUTER READABLE FORM

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

Dear Sir:

Applicants request that the previously submitted sequence information filed in computer readable form in Application Number 08/212,185 be used in the present Application. This sequence information was the only computer readable form filed in that Application.

Respectfully submitted,



MICHAEL D. DAVIS
Attorney for Applicant(s)
Registration No. 39,161

KLAUBER & JACKSON
411 Hackensack Avenue
Hackensack, New Jersey 07601
(201) 487-5800

Date: January 20, 2000